



INSTITUTO SUPERIOR DE CIÊNCIAS DA SAÚDE EGAS MONIZ

MESTRADO EM ANÁLISES CLÍNICAS

A HEMOGLOBINA FETAL COMO FATOR MODIFICADOR DA GRAVIDADE CLÍNICA DA ANEMIA DAS CÉLULAS FALCIFORMES: O EXEMPLO DOS ACIDENTES VASCULARES CEREBRAIS

Trabalho submetido por
Marta Isabel Benedito Pereira Nicolau
para a obtenção do grau de Mestre em Análises Clínicas

Maio de 2017



INSTITUTO SUPERIOR DE CIÊNCIAS DA SAÚDE EGAS MONIZ

MESTRADO EM ANÁLISES CLÍNICAS

A HEMOGLOBINA FETAL COMO FATOR MODIFICADOR DA GRAVIDADE CLÍNICA DA ANEMIA DAS CÉLULAS FALCIFORMES: O EXEMPLO DOS ACIDENTES VASCULARES CEREBRAIS

Trabalho submetido por
Marta Isabel Benedito Pereira Nicolau
para a obtenção do grau de Mestre em Análises Clínicas

Trabalho orientado por
Doutora Paula Faustino, Instituto Nacional de Saúde Doutor Ricardo Jorge

e coorientado por
**Prof. Doutora Manuela Caniça, Instituto Superior de Ciências da Saúde Egas
Moniz**

Maio de 2017

*Ao Miguel e à Madalena,
minhas fontes de inspiração
para esta caminhada...*

AGRADECIMENTOS

Ao Conselho Diretivo do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA), Dr. Fernando de Almeida e Eng.º José Maria Albuquerque, entidade na qual pude desenvolver o presente trabalho.

À minha orientadora de tese, Doutora Paula Faustino, responsável pelo Grupo de Investigação e Desenvolvimento (I&D) em Hemoglobinopatias, metabolismo do ferro e patologias associadas do INSA, pessoa a quem devo o tema deste projeto, uma disponibilidade e amizade ímpares.

À Professora Doutora Manuela Caniça, coorientadora e coordenadora do mestrado em Análises Clínicas do Instituto Superior de Ciências da Saúde Egas Moniz, pelos ensinamentos e disponibilidade demonstrada.

À Dr.ª Glória Isidro, coordenadora do Departamento de Genética Humana do INSA e ao Doutor João Lavinha, coordenador da Unidade de I&D do Departamento de Genética Humana do INSA pelo acolhimento no Departamento/Unidade de I&D.

Ao Doutor Luís Vieira, coordenador da Unidade de Tecnologia e Inovação do Departamento de Genética Humana do INSA e à Joana Mendonça pela realização e análise das amostras sequenciadas por NGS.

À Dr.ª Armandina Miranda e à Filomena Seuanes, do Departamento de Promoção da Saúde e Prevenção das Doenças não Transmissíveis do INSA, pela transmissão de conhecimentos na área de Hematologia.

À Doutora Marta Barreto, do Departamento de Epidemiologia do INSA, pela preciosa ajuda no tratamento de dados no SPSS.

Aos colegas Marisa Silva, Ildegário Semente e Daniela Brás do Grupo de I&D em Hemoglobinopatias, metabolismo do ferro e patologias associadas do INSA pelo apoio nos trabalhos realizados.

Ao Dr. Armindo Gonçalves e ao Dr. Fernando Fernandes, do Laboratório de Análises Clínicas de Beja (LACLIBE), sem a permissão dos quais não teria sido possível desenvolver este projeto.

A todos os meus colegas do LACLIBE, que das mais variadas formas me apoiaram durante este período.

À Ana Morgado, Sandra Viegas e Daniela de Almeida, pela amizade e apoio nas mais diversas fases deste trabalho.

Às minhas colegas de Mestrado, assim como aos professores e colaboradores do Instituto Superior de Ciências da Saúde Egas Moniz.

Ao meu irmão Nuno, minha cunhada Rita e meus sobrinhos, Francisco e João, pela disponibilidade, hospitalidade e amizade.

Aos meus pais, Faustino e Bernardina Nicolau, pelo apoio que sempre me deram na aquisição de conhecimentos ao longo de todo o meu percurso académico.

Ao José Miguel Almeida, meu marido e aos meus filhos Miguel e Madalena, que sempre me acompanharam e apoiaram...

A Todos Muito Obrigada!

RESUMO

A anemia das células falciformes (SCA) é uma anemia hemolítica hereditária de transmissão autossômica recessiva devida à mutação HBB:c.20A>T no gene da β -globina que dá origem à hemoglobina S (HbS). A HbS quando em desoxigenação polimeriza no glóbulo vermelho, conferindo-lhe a forma de foice e tornando-o rígido e frágil. A SCA caracteriza-se por manifestações clínicas de gravidade variável que podem incluir, por ex., hemólise crônica, crises vaso-oclusivas e acidente vascular cerebral (AVC). A variabilidade fenotípica é devida a fatores ambientais e genéticos, e o nível de hemoglobina fetal (Hb F) é considerado o principal modulador da doença.

Neste trabalho pretendeu-se investigar em crianças com SCA se o nível de HbF é modulado pelo seus genótipos em regiões polimórficas conhecidas nos genes *HBG2*, *BCL11A*, *HBSIL-MYB* e *KLF1* e, por outro lado, se o nível de HbF está relacionado com o grau de hemólise que apresentam e o risco de desenvolvimento de AVC.

Foram estudadas 67 crianças com SCA de origem Africana. Estas foram agrupadas de acordo com o grau de vasculopatia cerebral (Controlo, Risco e AVC). Para o estudo molecular foi usada a PCR-RFLP, sequenciação de Sanger e sequenciação de nova geração. Efetuaram-se estudos *in silico* e análise estatística em SPSS.

Os resultados permitiram concluir que um nível baixo de HbF é um fator de risco para o desenvolvimento de AVC ($p=0,005$). Observou-se também que os doentes com os genótipos mais raros em *HBG2* (rs7482144_TT+TC) apresentam níveis mais elevados de HbF ($p=0,031$). No gene *BCL11A* concluiu-se que o genótipo rs11886868_CC e o rs4671393_alelo_A apresentam-se tendencialmente associados a níveis mais elevados de HbF e, pelo contrário, os polimorfismos estudados em *HBSIL-MYB* não revelaram associação com os níveis de HbF. A análise do gene *KLF1* revelou que 82,8% dos doentes possuía pelo menos uma variante neste gene, foram detetadas 11 variantes diferentes sendo uma delas não descrita, a Q342H, mas de efeito benigno na proteína.

Concluiu-se que é de grande interesse considerar o efeito conjunto das referidas variantes nos genes analisados uma vez que a combinação no indivíduo das variantes mais raras *versus* as mais comuns está relacionada com os níveis de HbF ($p=0,021$) constituindo portanto um perfil genético de valor prognóstico. Os fatores genéticos analisados mostraram ser importantes modificadores dos níveis de HbF e poderão vir a constituir alvos de abordagem terapêutica.

Palavras-chave: Anemia das células falciformes, hemoglobina fetal, AVC.

ABSTRACT

Sickle Cell Anemia (SCA) is a hereditary hemolytic anemia with autosomal recessive transmission caused by the HBB:c.20A>T mutation in the β -globin gene, giving rise to hemoglobin S (HbS). Deoxygenated HbS polymerizes in the red cell, which acquires a sickle shape and becomes rigid and fragile. SCA is characterized by a variable clinical severity, including chronic hemolysis, vaso-occlusive episodes, and ultimately stroke. This phenotypic variability is due to environmental and genetic factors, and level of fetal hemoglobin (Hb F) has been proposed as the more important disease modulator.

In this work, we aimed to investigate if the level of HbF in children with SCA is modulated by their genotypes in known polymorphic regions in the genes *HBG2*, *BCL11A*, *HBSIL-MYB* and *KLF1*. We also aimed to investigate whether the level of HbF in these children is related with the degree of hemolysis and the risk of stroke.

We have studied 67 children with SCA, of African origin. These children were grouped in accordance with their degree of cerebral vasculopathy (Control, Risk and Stroke). For the molecular analysis, methodologies such as PCR-RFLP, Sanger sequencing and new generation sequencing were used. *In silico* studies were performed with *PolyPhen-2* software and the statistical analysis was done on SPSS software.

The results obtained allowed to conclude that a low HbF level is a risk factor for stroke development ($p=0.005$). It was also observed that the patients with the rarest genotypes in *HBG2* (rs7482144_TT+TC) present higher levels of HbF ($p=0.031$). It was concluded that, in *BCL11A* gene, the rs11886868_CC genotype and the rs4671393_allele_A seemed to predispose to higher levels of HbF. On the other hand, the polymorphisms studied in *HBSIL-MYB* did not reveal an association with HbF levels. The analysis of *KLF1* gene revealed that 82.8% of the patients presented at least one variant in this gene. In this screening, there were detected 11 different variants, being one of them novel - the Q342H - whose effect at the protein level should be benign.

We conclude that it is of great interest to consider the whole effect of the variants in the genes analyzed, since at the individual level, the combination of the rarest variants *versus* the more common is related with the HbF levels ($p=0.021$), constituting a genetic pattern with prognostic value. The genetic factors analyzed in this work have shown to be important modifiers of HbF levels and may become therapeutic targets.

Keywords: Sick Cell Anemia, fetal hemoglobin, stroke.

ÍNDICE

CAPITULO I	11
INTRODUÇÃO	11
I.1. Hemoglobina Humana.....	13
I.1.1. Estrutura.....	13
I.1.2. Estrutura dos agrupamentos génico das globinas (α e β).....	14
I.1.3. Eritropoiese durante as fases embrionária, fetal e adulta.....	15
I.2. Patologias que envolvem a hemoglobina – as hemoglobinopatias.....	17
I.3. Anemia das células falciformes (Drepanocitose).....	18
I.3.1. Origem genética e prevalência.....	18
I.3.2. Base genética	19
I.3.3. Transmissão da doença	20
I.3.4. Manifestações clínicas.....	21
I.4. Fatores modificadores do fenótipo da drepanocitose.....	22
I.4.1. Hemoglobina fetal e estratégias terapêuticas baseadas no seu aumento	22
I.4.2. Gene <i>HBG2</i> (cromossoma 11p).....	25
I.4.3. Gene <i>BCL11A</i> (cromossoma 2p).....	25
I.4.4. Gene <i>HBS1L-MYB</i> (cromossoma 6p)	26
I.4.5. Gene <i>KLF1</i> (cromossoma 19p).....	27
I.5. As vasculopatias cerebrais nas crianças com drepanocitose.....	28
CAPITULO II	33
OBJETIVOS	33
CAPITULO III	39
MATERIAL E MÉTODOS	39
III.1. Amostra	41
III.1.1. Amostra populacional	41
III.1.2. Amostra biológica.....	42
III.2. Métodos	42
III.2.1. Amplificação de DNA genómico pela técnica de <i>Polymerase Chain Reaction</i> (PCR).....	43
III.2.2. <i>Restriction Fragment Length Polymorphism</i> (RFLP).....	43
III.2.3. Sequenciação automática pelo método de Sanger.....	44
III.2.4. PCR-Longo e Sequenciação de Nova Geração (<i>Next Generation Sequencing</i> , NGS).....	45
III.3. Análise Estatística por SPSS.....	46
CAPITULO IV	49

RESULTADOS E DISCUSSÃO	49
IV.1. Estudo dos parâmetros hematológicos e de hemólise na população em estudo	51
IV.2. Pesquisa molecular de mutação drepanocítica	52
IV.3. Caracterização molecular de regiões polimórficas reconhecidas como potencialmente modificadoras	54
IV.3.1. <i>HBG2</i> : SNP rs7482144	54
IV.3.2. <i>BCL11A</i> : SNPs rs11886868 e rs4671393	55
IV.3.4. <i>HBS1L-MYB</i> : SNPs rs4895441; rs9399137 e rs9389268	58
IV.4. Estudos de associação entre os genótipos e os diferentes sub-fenótipos da doença	63
IV.4.1. Variantes genéticas versus nível de hemoglobina fetal	63
IV.4.2. Variantes genéticas versus risco de AVC	66
IV.4.3. Estudos de associação entre o nível de Hb e o risco de AVC	67
IV.4.4. Estudos de associação entre o nível de Hb F e o risco de AVC	68
IV.5. Pesquisa e caracterização de variantes no gene <i>KLF1</i>	69
IV.5.1. Amplificação do gene <i>KLF1</i> através da técnica de PCR-longo	69
IV.5.2. Análise bioinformática dos resultados de NGS e validação dos resultados	70
IV.5.3. Vantagens e desvantagens da utilização da metodologia de NGS	72
IV.5.4. Tipo de variantes encontradas, consequências funcionais e frequências alélicas	73
IV.5.5. Estudo de associação das variantes no gene <i>KLF1</i> com o nível de HbF	76
IV.5.6. Análise <i>in silico</i> através do software <i>PolyPhen-2</i> , das consequências na expressão gênica da nova variante identificada	77
IV.6. As diferenças nas frequências das variantes no gene <i>KLF1</i> entre populações e patologias	78
IV.7. A ação de efeito aditivo das variantes sobre o nível de HbF	79
IV.8. Estudos de associação entre os genótipos das variantes do <i>KLF1</i> e o risco de AVC	81
CAPITULO V	85
CONCLUSÕES E PERSPETIVAS FUTURAS	85
CAPITULO VI	91
BIBLIOGRAFIA	91
ANEXOS	

ÍNDICE DE FIGURAS

Figura I.1: Estrutura da hemoglobina.....	13
Figura I.2: Organização cromossômica dos <i>clusters</i> α e β -globina	15
Figura I.3: Regulação molecular do <i>switch</i> de hemoglobina fetal para hemoglobina adulta.....	16
Figura I.4: Esfregaço sanguíneo de um doente com anemia das células falciformes.....	19
Figura I.5: Novos alvos para a ativação da transcrição do gene da γ -globina.....	28
Figura I.6: Vasculopatia e AVC na anemia das células falciformes	30
Figura IV.1: Pesquisa de mutação drepanocítica por PCR-RFLP.....	53
Figura IV.2: Resultado do produto de PCR amplificado do gene <i>HBG2</i> para pesquisa do SNP rs7482144.....	54
Figura IV.3: Resultado da digestão enzimática do fragmento de PCR amplificado do gene <i>HBG2</i> com a enzima de restrição <i>XmnI</i> para pesquisa do SNP rs7482144	55
Figura IV.4: Resultado do produto de PCR amplificado do gene <i>BCL11A</i> para pesquisa do SNP rs11886868.....	56
Figura IV.5: Resultado da digestão enzimática do fragmento de PCR amplificado do gene <i>BCL11A</i> com a enzima de restrição <i>MbolI</i> para pesquisa do SNP rs11886868.....	56
Figura IV.6: Resultado do produto de PCR amplificado do gene <i>BCL11A</i> para pesquisa do SNP rs4671393.....	57
Figura IV.7: Resultado da sequenciação do produto de PCR amplificado do gene <i>BCL11A</i> para pesquisa do SNP	58
Figura IV.8: Resultado do produto de PCR amplificado da região intergênica dos genes <i>HBS1L-MYB</i> para pesquisa do SNP rs4895441	59
Figura IV.9: Resultado da digestão enzimática do fragmento de PCR amplificado da região intergênica dos genes <i>HBS1L-MYB</i> com a enzima de restrição <i>RsaI</i> para pesquisa do SNP rs4895441.....	59
Figura IV.10: Resultado do produto de PCR amplificado do gene <i>MYB</i> para pesquisa do SNP rs9399137.....	60
Figura IV.11: Resultado da sequenciação do produto de PCR amplificado do gene <i>MYB</i> para pesquisa do SNP rs9399137	61
Figura IV.12: Resultado do produto de PCR amplificado do gene <i>MYB</i> para pesquisa do SNP rs9389368.....	62
Figura IV.13: Resultado da sequenciação do produto de PCR amplificado do gene <i>MYB</i> para pesquisa do SNP rs9389268	62
Figura IV.14: <i>Box-Plot</i> dos genótipos do rs11886868 e rs4671393 do gene <i>BCL11A</i> em função dos valores de HbF na população em estudo	63
Figura IV.15: <i>Box-Plot</i> dos genótipos do rs7482144 do gene <i>HBG2</i> em função dos valores de HbF na população em estudo	65

Figura IV.16: <i>Box-Plot</i> dos níveis de Hb (g/dL) para os três grupos em estudo, o grupo controle, o grupo de risco e o grupo AVC.....	67
Figura IV.17: <i>Box-Plot</i> dos valores de HbF (%) para os três grupos em estudo, o grupo controle, o grupo de risco e o grupo AVC.....	68
Figura IV.18: Resultado do produto de PCR amplificado do gene <i>KLF1</i> para posterior sequenciação por NGS	70
Figura IV.19: Resultado da sequenciação de Sanger que permitiu confirmar as variantes encontradas por NGS na população em estudo, no gene <i>KLF1</i>	71
Figura IV.20: Representação esquemática das mutações encontradas na população em estudo no gene <i>KLF1</i>	72
Figura IV.21: Alinhamento da sequência peptídica da proteína KLF1	77
Figura IV.22: Resultado da análise do modelo <i>HumDiv</i> e <i>HumVar</i> através do programa <i>PolyPhen-2</i>	78
Figura IV.23: <i>Box-Plot</i> das amostras para a combinação das variantes para o gene <i>KLF1</i> , <i>HBG2</i> (<i>XmnI</i> ; rs7482144), <i>BCL11A</i> (rs4671393), <i>HBS1L-MYB</i> (rs4895441) e <i>MYB</i> (rs9399137).....	80

ÍNDICE DE TABELAS

Tabela III.1. – Resumo das metodologias moleculares utilizadas no estudo de cada variante genética	42
Tabela IV.1: Parâmetros hematológicos e de hemólise da população em estudo	52
Tabela IV.2: Frequências genotípicas e alélicas para o SNP <i>HBG2</i> rs7482144.....	55
Tabela IV.3: Frequências genotípicas e alélicas para o SNP <i>BCL11A</i> rs11886868.....	57
Tabela IV.4: Frequências genotípicas e alélicas para o SNP <i>BCL11A</i> rs4671393.....	58
Tabela IV.5: Frequências genotípicas e alélicas para o SNP <i>HBS1L_MYB</i> rs4895441	60
Tabela IV.6: Frequências genotípicas e alélicas para o SNP <i>MYB</i> rs9399137	61
Tabela IV.7: Frequências genotípicas e alélicas para o SNP <i>MYB</i> rs9389268	62
Tabela IV.8: Estudo estatístico para os vários SNPs mostrando as várias combinações dos genótipos e dos alelos com os vários grupos estudados	66
Tabela IV.9: Variantes identificadas no gene <i>KLF1</i> , através de NGS.....	70
Tabela IV.10: Frequência alélica das várias variantes identificadas no gene <i>KLF1</i> através de NGS	73
Tabela IV.11: Mutações <i>missense</i> no exão 2 do gene <i>KLF1</i> e respectivos valores de HbF	76
Tabela IV.12: Níveis de HbF dos indivíduos com anemia das células falciformes, de acordo com a presença de variantes em <i>KLF1</i> e outros polimorfismos genéticos nos genes <i>HBG2</i> , <i>BCL11A</i> , <i>HBS1L-MYB</i> e <i>MYB</i>	80
Tabela IV.13: Estudo estatístico para as diferentes variantes genéticas identificadas no gene <i>KLF1</i> , mostrando as várias combinações dos genótipos com os vários grupos estudados	82

ABREVIATURAS

A – Adenina

AVC – Acidente Vascular Cerebral

BCL11A - *B-cell CLL/lymphoma 11A*

C – Citosina

CSSCD – Estudo Cooperativo da Doença das Células Falciformes (*Cooperative Study of Sick Cell Disease*)

ddNTP - didesoxiribonucleótidos

DNA – Ácido Desoxiribonucleico

dNTP – desoxiribonucleótidos

ET-1 – Endotelina-1

EtBr – Brometo de Etídio

FAO - *Food and Agriculture Organization*

G – Guanina

Glu – Ácido Glutâmico

Hb – Hemoglobina

HbA – Hemoglobina Adulta

HBB – Gene da β -globina

HbC – Hemoglobina C

HbD – Hemoglobina D

HbF – Hemoglobina Fetal

HBG2 - Gene da γ -globina

HbS – Hemoglobina S

HBS1L - *HBS1 like translational GTPase*

HDE – Hemólise Disfunção Endotelial

HPFH – Persistência Hereditária de Hemoglobina Fetal (*Hereditary Persistence of Fetal Hemoglobin*)

HPLC – Cromatografia Líquida de Alta Resolução

HS – Sítio Hipersensível a DNase I

HWE – Equilíbrio de Hardy-Weinberg

INSA – Instituto Nacional de Saúde Doutor Ricardo Jorge

Kb - Kilobase

KLF1 – *Kruppel like factor 1*

LCR – *Locus Control Region*

LDH – Lactato Desidrogenase

MCS – Sequências Conservadas Multiespécies (*Multispecies Conserved Sequences*)

MCV – Volume Corpuscular Médio

MYB - *proto-oncogene transcription factor*

NGS – Sequenciação de Nova Geração (*Next Generation Sequencing*)

NO – Óxido Nítrico

OMS – Organização Mundial de Saúde

pb – pares de bases

PCR – Reação em cadeia da polimerase (*Polymerase Chain Reaction*)

RBC – Glóbulos vermelhos ou eritrócitos (*Red Blood Cells*)

RDW - *Red Cell Distribution Width*

RFLP – *Restriction Fragment Length Polymorphism*

SCA – Anemia das células falciformes (*Sickle Cell Anemia*)

SNP – *Single Nucleotide Polymorphism*

T – Timina

Val – Valina

VVO – Viscoso Vaso-Oclusivo

CAPITULO I
INTRODUÇÃO

I – INTRODUÇÃO

I.1. Hemoglobina Humana

I.1.1. Estrutura

Os eritrócitos, também chamados glóbulos vermelhos (RBC – *Red Blood Cells*), são as células do organismo que transportam o oxigénio dos pulmões para os tecidos. Os eritrócitos são de forma bicôncava, o que aumenta a área de superfície da célula e facilita a difusão de oxigénio e dióxido de carbono. No entanto, são muito flexíveis e maleáveis o que lhes permite fluir através dos capilares. Os eritrócitos imaturos, chamados reticulócitos, normalmente representam 1-2% dos glóbulos vermelhos no sangue. Em ambos, a cor avermelhada é devida à presença de hemoglobina (Theml, Diem & Haferlach, 2004).

As moléculas de hemoglobina (Hb) humana são um conjunto de proteínas compostas por 4 sub-unidades, cada uma contendo um grupo prostético (heme) e uma cadeia polipeptídica, denominada globina. Cada molécula de hemoglobina contém duas globinas do tipo alfa e duas do tipo beta (Figura I.1). O grupo heme é formado por um complexo de protoporfirina e ferro e é através deste que a Hb liga o oxigénio (Lettre & Bauer, 2016; Schechter, 2008).

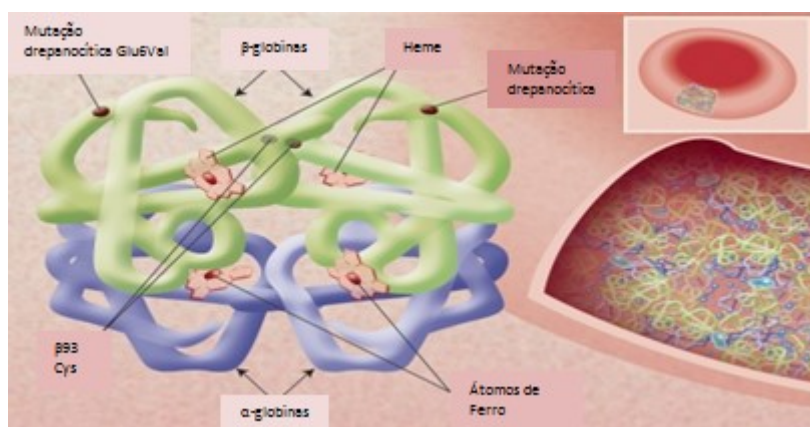


Figura I.1: Estrutura da hemoglobina. A figura mostra a estrutura da molécula de hemoglobina com as duas cadeias de globina alfa e beta e os 4 grupos heme com os átomos de ferro. Também se mostra o local da mutação que dá origem à hemoglobina S na anemia das células falciformes. À direita está representado um eritrócito maduro (Adaptado de Schechter, 2008).

I.1.2. Estrutura dos agrupamentos génico das globinas (α e β)

Os genes humanos que codificam as cadeias de globina localizam-se em dois agrupamentos génicos (*clusters*): os do tipo α (ξ , $\alpha 2$ e $\alpha 1$) no cromossoma 16 e os do tipo β (ϵ , $^G\gamma$, $^A\gamma$, δ e β) no cromossoma 11. Em ambos os *clusters* os genes são dispostos na mesma orientação transcricional e na mesma ordem em que serão expressos durante o desenvolvimento (Figura I.2) (Stamatoyannopoulos, 2005; Saki *et al.*, 2016).

Em indivíduos saudáveis, a síntese das cadeias α - e β -globina é finamente equilibrada durante a diferenciação eritróide terminal e, embora cada cópia do cromossoma 16 possua dois genes α idênticos e cada cópia do cromossoma 11 possua apenas um gene β , há sempre uma produção equimolar das cadeias de globina tipo α e β (Frenette & Atweh, 2007).

A expressão coordenada dos genes em cada *cluster* nos vários estádios de desenvolvimento depende, em parte, de elementos reguladores críticos localizados a montante dos genes. O *cluster* α -globina humano situa-se a cerca de 150 kb do telómero no braço curto do cromossoma 16 e está rodeado por genes amplamente expressos, numa região rica em GC. A cerca de 25 a 65 kb a montante dos genes de α -globina existem quatro sequências conservadas multiespécies (MCS- *Multispecies Conserved Sequences*), MCS-R1 a MCS-R4, que parecem estar envolvidas na regulação deste *cluster* (Razin, *et al.*, 2012). Três destes elementos, MCS-R1 a -R3 encontram-se dentro de intrões do gene *NPR3L*, um gene *housekeeping* transcrito em direção oposta à transcrição dos genes de α -globina (Razin, *et al.*, 2012).

No *cluster* β , existem cinco sítios hipersensíveis à DNase I (HS-1 a -5) agrupados para formar o LCR (*Locus Control Region*). Estes elementos são necessários para a manutenção de uma configuração de cromatina aberta do *locus*, para permitir o acesso dos fatores de transcrição aos elementos reguladores que medeiam a expressão de cada gene neste *cluster*. Os locais HS-2 a -5 são considerados os subdomínios de LCR; HS-2 a -4 são estimuladores eritróides específicos e contêm os locais de ligação para os diferentes fatores de transcrição; HS-2 é o principal estimulador porque contém vários locais de ligação aos fatores de transcrição (Figura I.2) (Kiefer, Hou, Little & Dean, 2008; Razin, *et al.*, 2012).

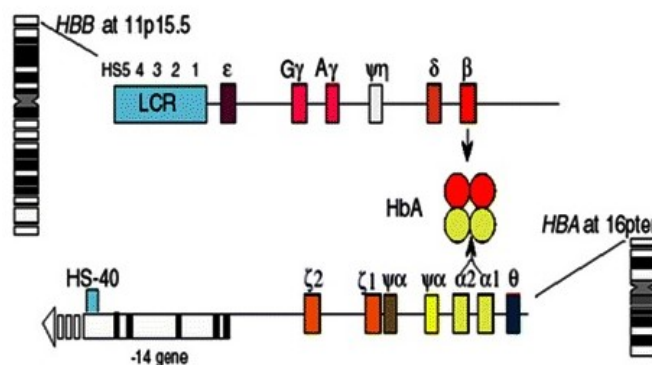


Figura I.2: Organização cromossômica dos *clusters* α e β-globina. Os genes do *cluster* da β-globina (ε, $G\gamma$, $A\gamma$, δ e β) estão localizados no cromossoma 11 na mesma ordem em que são expressos durante o desenvolvimento. Os genes do *cluster* da α-globina (ζ, α₁ e α₂) estão localizados no cromossoma 16 na mesma ordem em que são expressos durante o desenvolvimento. O LCR e o HS-40 são os principais elementos regulatórios localizados a montante, respectivamente, do *cluster* da β e α-globina, que são necessários para o elevado nível de expressão destes genes (Adaptado de Ngo & Steinberg, 2015).

I.1.3. Eritropoiese durante as fases embrionária, fetal e adulta

Os genes das globinas humanas foram dos primeiros genes a serem clonados na década de 70, e de seguida a estrutura do conjunto da β-globina foi mapeada e sequenciada. Nesta altura, observou-se que os genes da γ-globina no *cluster* da β-globina eram silenciados após sofrerem metilação do DNA nas células eritróides adultas, e o mesmo não se verificava nas células eritróides embrionárias (Sankaran, 2011).

O agrupamento génico da β-globina localiza-se no cromossoma 11 e inclui 5 genes do tipo β: ε-globina (*HBE*, expresso principalmente no 1º trimestre de gestação); $G\gamma$ -globina e $A\gamma$ -globina (*HBG2* e *HBG1*; ambos expressos durante a gestação, especialmente no 2º e 3º trimestre); δ-globina e β-globina (*HBD* e *HBB*; os genes *minor* e *major* expressos depois do nascimento) (Bauer, Kamran & Orkin, 2012). Assim, no início do primeiro trimestre de gestação, existe uma expressão robusta de uma forma embrionária de uma globina do tipo β, conhecida como ε-globina dentro da posição primitiva dos eritrócitos no saco embrionário. Pouco tempo depois, quando a produção dos primeiros eritrócitos anucleados definitivos começa a partir de células progenitoras no fígado fetal, a molécula de globina produzida é a γ-globina codificada pelos genes $G\gamma$ -globina e $A\gamma$ -globina (Figura I.3). Duas cadeias de γ-globina combinam-se com duas cadeias de α-globina resultando um tetrâmero (α₂γ₂) de hemoglobina fetal (HbF). Esta continua sendo a hemoglobina predominante durante grande parte da gestação. Pouco tempo após

nascimento, há uma mudança da expressão predominante de HbF para a hemoglobina adulta (HbA), que é mediada por um *switch* transcricional em progenitores eritróides definitivos de γ - para β -globina (Figura I.3) (Sankaran & Orkin, 2013). Assim, A HbF é predominante durante a vida fetal e nos primeiros meses após o nascimento, depois decresce gradualmente até cerca de 1% no período adulto, sendo substituída por 97% de HbA que contém duas cadeias α -globina e duas cadeias β -globina ($\alpha_2 \beta_2$) e 2-3% de HbA₂ constituída por duas cadeias α -globina e duas cadeias δ -globina ($\alpha_2 \delta_2$) (Costa, Capuano, Sommese & Napoli, 2015; Lettre & Bauer, 2016).

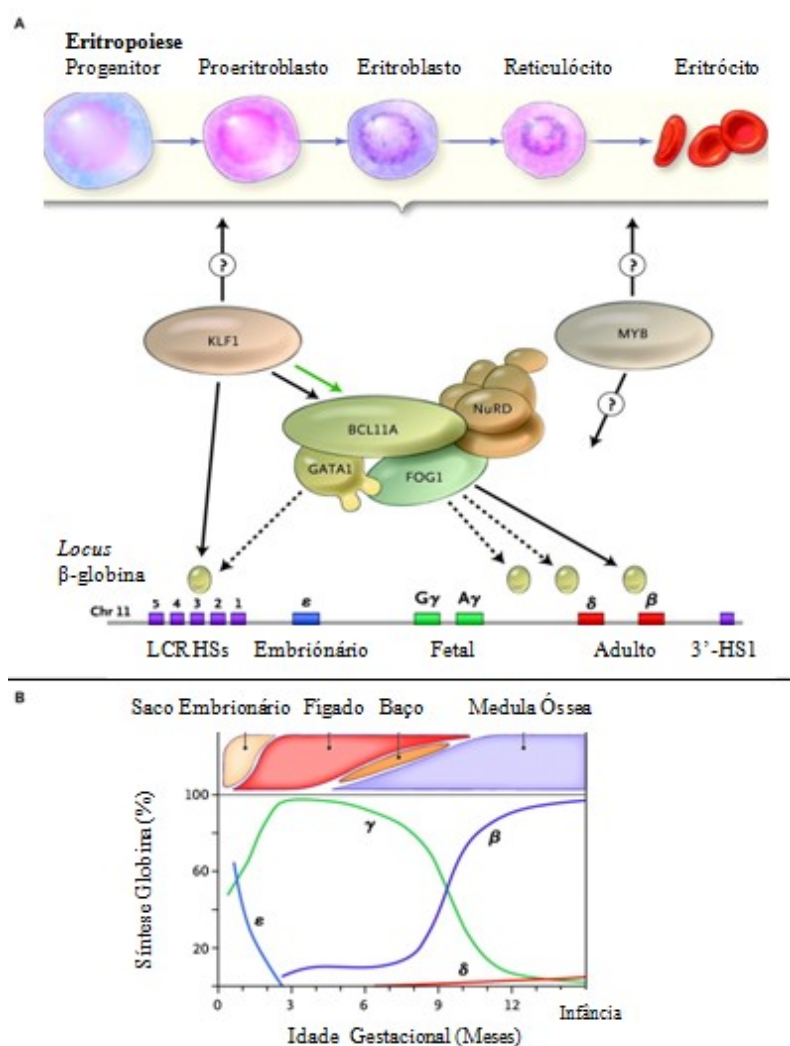


Figura I.3: Regulação molecular do *switch* de hemoglobina fetal para hemoglobina adulta. **A** – A figura mostra as várias fases da eritropoiese desde a célula progenitora até ao eritrócito maduro. O agrupamento génico da β -globina humana localiza-se no cromossoma 11 e é constituído por um conjunto de 5 genes que se expressam em determinada fase da vida e uma região regulatória distal chamada LCR. A figura também mostra fatores transcricionais que interatuam com o LCR e com os vários genes no *cluster*, tais como o *KLF1*, *BCL11A* e *MYB*. **B** – Síntese das várias globinas do tipo β deste a fase embrionária até ao nascimento, evidenciando o *switch* da expressão do gene da β -globina por volta do nascimento (Adaptado de Sankaran, 2011).

A expressão diferencial dos genes da globina ao longo do desenvolvimento é o resultado de um processo complexo e ainda em parte desconhecido. No entanto, sabe-se que o LCR, juntamente com vários fatores transcricionais, tanto ativadores como repressores, interagem diretamente com os diferentes genes no *cluster*, de uma forma sequencial, a partir da região 5' do gene ϵ -globina no embrião até à região 3' dos genes δ e β -globina nos adultos. De maneira semelhante, no agrupamento génico da α -globina a comutação na expressão dos genes de ξ - para α -globina é controlada, predominantemente, ao nível da transcrição, embora também envolvam mecanismos pós-transcricionais (Stamatoyannopoulos, 2005).

I.2. Patologias que envolvem a hemoglobina – as hemoglobinopatias

As patologias que envolvem a Hb denominam-se hemoglobinopatias e são as patologias genéticas recessivas mais comuns nas populações humanas (Bauer, Kamran & Orkin, 2012). Estas patologias são doenças Mendelianas que afetam apenas um único gene, mas com um fenótipo bastante heterogéneo. Na parte proteica da hemoglobina, a globina, estão descritas mais de 1500 mutações pontuais que alteram a sua estrutura, função e expressão (Ngo & Steinberg, 2015). Estima-se que o número de crianças que nascem por ano com hemoglobinopatias de significado clínico seja de 300000 a 500000, e, destes, 70% são doentes com Anemia das Células Falciformes (SCA) (Ngo & Steinberg, 2015; Paciaroni, *et al.*, 2014). A sua prevalência altera-se consoante a localização geográfica e a origem étnica.

As hemoglobinopatias podem ser do tipo qualitativo ou quantitativo. Nas hemoglobinopatias qualitativas encontram-se as variantes de hemoglobina, são resultantes da alteração da estrutura de uma cadeia de globina, e afetam as cadeias α ou β (por exemplo, HbS, HbC, HbD, etc). As hemoglobinopatias do tipo quantitativo são denominadas talassémias, são resultantes da ausência ou diminuição de síntese de uma cadeia de globina (por exemplo, beta-talassémia, alfa-talassémia, persistência hereditária de hemoglobina fetal – HPFH) (Danjou, *et al.*, 2012).

A beta-talassémia surge quando há uma redução ou ausência da síntese de cadeias β -globina, resultando num excesso de cadeias α não ligadas, as quais precipitam nos precursores eritróides da medula óssea conduzindo à morte prematura da célula e a uma eritropoiese ineficaz (Danjou, *et al.*, 2012; Danjou, Anni & Galanello, 2011; Finotti, *et al.*, 2015). Estes doentes têm anemia microcítica e hipocrómica, apresentam

esplenomegalia e hepatomegalia, deformações ósseas e sobrecarga de ferro. O tratamento (sintomático) passa por transfusões sanguíneas e quelatação do ferro ou, eventualmente, transplante de medula óssea (Danjou, Anni & Galanello, 2011; Paciaroni, *et al.*, 2014).

A alfa-talassémia é causada por mutações nos genes do *cluster* da α -globina (*HBA2* e *HBA1*) e é uma das doenças genéticas monogénicas mais comum a nível mundial. Assim, estima-se que 5% da população mundial seja portadora de alguma lesão molecular neste *cluster*. A causa mais frequente são as deleções de uma ou duas cópias do gene no mesmo cromossoma (Scheps, *et al.*, 2015).

As beta-talassémias e a SCA, são as hemoglobinopatias com maior mortalidade e morbidade em todo o mundo (Bauer, Kamran & Orkin, 2012).

A HPFH é uma condição clínica benigna caracterizada pela síntese contínua e aumentada de HbF no período adulto. Uma das razões para esse fato consiste na existência de alterações no promotor do gene da γ -globina (Lettre & Bauer, 2016) que impedem o seu silenciamento no período adulto. Bordin e seus colaboradores descreveram, pela primeira vez, polimorfismos de um nucleótido na região promotora do gene da γ -globina numa família onde segregava HPFH (Bordin, *et al.*, 1998).

I.3. Anemia das células falciformes (Drepanocitose)

I.3.1. Origem genética e prevalência

A anemia das células falciformes (SCA) ou drepanocitose foi inicialmente descrita em 1910 num estudante de medicina dentária com sintomas pulmonares, no qual foram identificados eritrócitos em forma de foice (Figura I.4). No entanto, não se sabia se esta alteração no sangue era uma doença ou uma manifestação de outra doença. Durante os 15 anos seguintes foram descritos vários casos semelhantes consolidando a ideia de que seria uma nova patologia, mas foi só em 1945 que Linus Pauling colocou a hipótese de ser uma doença com origem numa alteração na molécula de hemoglobina. Durante este ano mostrou-se ser uma doença autossómica recessiva. Ingram e seus colaboradores demonstraram em 1959 que esta hemoglobina alterada (hemoglobina S - HbS) se distinguia da hemoglobina normal (HbA) apenas num aminoácido. Em 1978 foi criado o estudo cooperativo da doença das células falciformes (CSSCD – *Cooperative Study of Sickle Cell Disease*), que realizou estudos *cohort* de forma a determinar a prevalência da

doença (Frenette & Atweh, 2007). Estima-se que cerca de 300000 crianças nascem por ano com SCA, maioritariamente em África (Bauer, Kamran & Orkin, 2012).

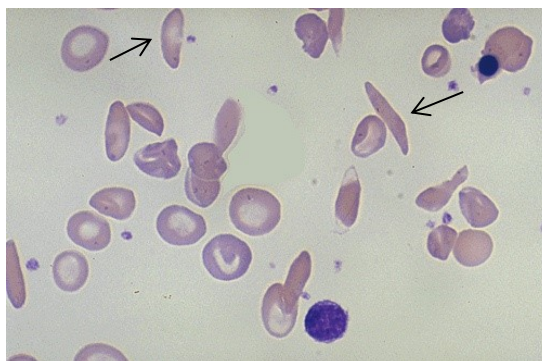


Figura I.4: Esfregaço sanguíneo de um doente com anemia das células falciformes. As setas apontam para os drepanócitos, ou também designados células em forma de foice (Adaptado de Theml, Diem & Haferlach, 2004).

I.3.2. Base genética

A anemia das células falciformes é uma anemia hemolítica causada por uma mutação no gene da β -globina (*HBB*) que altera a estrutura da hemoglobina dando origem à hemoglobina falciforme (HbS). A mutação caracteriza-se por uma alteração no codão 6 do gene da β -globina de uma adenina (A) para uma timina (T), traduzindo-se na substituição do ácido glutâmico em posição 6, por uma valina (Glu6Val), a qual não altera a capacidade da proteína se ligar ao oxigénio (Lettre & Bauer, 2016). Contudo, na presença de sangue pouco oxigenado a molécula de HbS, como tem apenas 1/5 da solubilidade da HbA, forma agregados de fibras poliméricas deformando os eritrócitos, conferindo-lhe a forma de foice (Figura I.4). Estes eritrócitos falciformes são menos deformáveis, causando obstrução vascular (vaso-oclusão) e isquemia local. A polimerização de moléculas de HbS também é acompanhada por danos na membrana e desidratação dos eritrócitos, aumentando a hemólise e originando anemia crónica.

Existem vários fatores que interferem, diretamente, na polimerização da HbS: i) uma diminuição do pH implica uma diminuição da afinidade da hemoglobina para o oxigénio, aumentando a polimerização da HbS; ii) o aumento da temperatura também aumenta a polimerização da HbS; iii) uma concentração intracelular de HbS mais elevada conduz a uma taxa aumentada de polimerização e iv) a presença de outras hemoglobinas limitam a polimerização da HbS. Sabe-se que a HbF e a HbA2 neutralizam o processo de polimerização mais eficazmente do que a HbA e a HbC (outra substituição do ácido

glutâmico na 6ª posição da cadeia de β -globina, mas por lisina) (Schnog, Duits, Luskiet, ten Cate & Rojer, 2004).

I.3.3. Transmissão da doença

A homozigotia para a mutação drepanocítica (HbS: $\alpha^A_2\beta^S_2$) é responsável pela forma mais grave da doença - SCA. As outras variantes genéticas da doença resultam da interação de diferentes mutações dos genes da β -globina humana, ou seja, são heterozigóticos compostos com uma mutação diferente no segundo alelo β (β^0 -talassémia, β^+ -talassémia) dando origem à talassodrepanocitose. Quando um gene β^S interage com um gene β -talassémia (um gene mutante da β -globina que não produz o mRNA normal da β -globina ou apresenta uma diminuição dos níveis), a gravidade do distúrbio falciforme resultante depende da gravidade da mutação de β -talassémia co-herdada. Quando o gene β -talassémia co-herdado é completamente inativo (β^0 -talassémia), o distúrbio falciforme resultante, conhecido como $S\beta^0$ -talassémia, tende a ser de gravidade semelhante ao homozigótico para HbS. Em contraste, quando o gene β -talassémia co-herdado é parcialmente ativo (β^+ -talassémia), o distúrbio falciforme resultante conhecido como $S\beta^+$ -talassémia pode ter um espectro de gravidade clínica variada. Se a mutação β^+ -talassémia for leve, como é comum em pessoas de ascendência africana, a $S\beta^+$ -talassémia resultante tende a ser clinicamente leve. Em contraste, se a mutação β^+ -talassémia for grave, como é comum no caso das populações do Mediterrâneo, o distúrbio falciforme tende a ser moderado a grave (Frenette & Atweh, 2007). Existem ainda casos em que o segundo alelo mutado também dá origem a uma variante de Hb, por exemplo HbC, HbD, etc. Quando o gene β^S interage com o gene β^C ou β^D , ou outra variante, dá origem aos chamados síndromes drepanocíticos (HbSC, HbSD, etc) os quais são em geral menos graves do que a SCA (HbSS).

Os indivíduos heterozigóticos para a variante da hemoglobina S (HbAS: [$\alpha^A_2\beta^A_2$, $\alpha^A_2\beta^S_2$] e [$\alpha^A_2\beta^A\beta^S$]) são clinicamente "normais", mas podem apresentar sintomas da anemia falciforme, isto é, em situações de extrema desoxigenação, como altitudes elevadas ou em grandes alterações físicas, os seus eritrócitos podem tornar-se falciformes (Schnog, *et al.*, 2004; Lettre & Bauer, 2016). A SCA é, portanto, uma doença recessiva nas suas manifestações clínicas, mas o gene afetado tem expressão dominante, uma vez que os indivíduos com HbAS podem apresentar eritrócitos falciformes em amostras de sangue desoxigenado (Schnog *et al.*, 2004).

Todas as formas principais de SCA apresentam anemia hemolítica, caracterizada por baixos níveis de hemoglobina, elevadas contagens de reticulócitos, níveis séricos de lactato desidrogenase (LDH) elevados, e baixos níveis de haptoglobina. O volume corpuscular médio (MCV) é, geralmente, normal. Os eritrócitos falciformes podem ser visualizados no esfregaço de sangue periférico (Figura I.4). A prova de solubilidade da HbS é realizada por turbidimetria, uma vez que a HbS é insolúvel na sua forma reduzida, quando na presença de tampão fosfato, devido à formação de polímeros que refratam e refletem a luz produzindo turvação. Esta prova confirma a presença de HbS mas não faz distinção entre os diferentes genótipos.

Os métodos para a determinação da presença de formas anormais de hemoglobina incluem a electroforese de hemoglobinas, a cromatografia líquida de alta resolução (HPLC), a focagem isoelétrica e a análise de mutações através da sequência de genes das cadeias de globina (Schnog, *et al.*, 2004)

I.3.4. Manifestações clínicas

Como já foi referido anteriormente, a SCA é uma doença autossómica recessiva causada por alteração num único gene, dando origem à HbS. No entanto, existe um amplo espectro de manifestações fenotípicas e complicações que faz com que esta doença se identifique com as doenças multigénicas (Driss, *et al.*, 2009; Ngo & Steinberg, 2015). As manifestações clínicas da anemia das células falciformes derivam, essencialmente, de dois fenómenos: hemólise e vaso-occlusão. Dado o seu espectro de prevalência e gravidade, foram feitas tentativas de categorizar doentes em subgrupos que permitiriam aos médicos antecipar complicações mais graves (Kato, Gladwin & Steinberg, 2007). Estes autores propuseram dois subfenótipos baseados nestas duas principais manifestações: o subfenótipo viscoso-vaso-oclusivo (VVO), com níveis de hemoglobina relativamente elevados e relacionado com a polimerização de HbS; e o subfenótipo hemólise-disfunção endotelial (HDE), associado a níveis baixos de hemoglobina e níveis elevados de marcadores hemolíticos (contagem de reticulócitos, lactato desidrogenase sérica, hemoglobina plasmática e bilirrubinas) (Kato, *et al.*, 2007).

Os episódios dolorosos agudos são os principais sintomas clínicos da SCA. Os doentes experimentam crises dolorosas devido ao enfarte da medula óssea, o que leva a isquemia tecidual, causando dor muito intensa, que pode requerer hospitalização e tratamento com opióides (Driss, *et al.*, 2009; Steinberg, 2009).

Existem estudos que revelam que, aproximadamente, 11% dos doentes vão desenvolver um acidente vascular cerebral (AVC) por volta dos 20 anos de idade, enquanto 24% por volta dos 45 anos. Assim, estes estudos também mostram que cerca de 50% destes doentes morre antes dos cinquenta anos de idade (Frenette & Atweh, 2007).

A fisiopatologia da hipercoaguabilidade na SCA é multifatorial e é o resultado da alteração da maioria dos componentes do sistema da hemostase. Estas alterações nas plaquetas e nos sistemas fibrinólicos pro-coagulantes e anticoagulantes são sobretudo de natureza pro-trombótica (Pakbaz & Wun, 2014). Vários investigadores mostraram a existência de biomarcadores da ativação da cascata da coagulação durante a crise vaso-oclusiva. Estes marcadores mostram o estado de hipercoaguabilidade na SCA. A trombina, os complexos trombina-antitrombina, os produtos do fibrinogénio e o D-dímero estão aumentados, o que está associado com a ativação do endotélio vascular na SCA juntamente com a hipertensão arterial pulmonar. O fator V e o fator VII estão diminuídos devido ao aumento da atividade do fator tecidual (Pakbaz & Wun, 2014).

I.4. Fatores modificadores do fenótipo da drepanocitose

I.4.1. Hemoglobina fetal e estratégias terapêuticas baseadas no seu aumento

Em 1948, Janet Watson verificou que recém-nascidos não manifestavam complicações associadas à SCA até cerca dos 6 meses de idade, concluindo que os eritrócitos destes estão protegidos em relação aos eritrócitos dos adultos. Assim postulou-se a hipótese de que esta proteção nas crianças seria devido à elevada concentração de HbF circulante nos eritrócitos. Alguns anos mais tarde outros autores descreveram doentes com SCA assintomáticos, mas que apresentavam simultaneamente a HPFH, condição caracterizada pela síntese contínua e aumentada de HbF no adulto (Lettre & Bauer, 2016).

Algumas populações de doentes com SCA provenientes da Arábia Saudita e da Índia apresentam níveis elevados de HbF, associados com a diminuição dos sintomas clínicos da doença (Sankaran, 2011). Em doentes com β -talassémia também se encontraram níveis elevados de HbF associados à diminuição dos sintomas e que se mantiveram até à idade da criança começar a diminuir os níveis de HbF (Galanello *et al*, 2009).

Em estudos realizados pela CSSCD os níveis elevados de HbF foram associados com o aumento da sobrevivência e à diminuição dos episódios de dor nos doentes com SCA

(Frenette & Atweh, 2007; Lettre & Bauer, 2016). Por outro lado, outras complicações como o acidente vascular cerebral (AVC) e o priapismo foram consideradas, por alguns autores, independentes dos níveis de HbF, o que se pode dever ao facto destes sintomas serem ocasionais e terem, consequentemente, um limite estatístico para se detetar o efeito (Lettre & Bauer, 2016).

Os tratamentos mais comuns da SCA e da β -talassémia envolvem transfusões de sangue regulares e agentes quelantes para remover o excesso de ferro transfundido. Outra terapia é o transplante alogénico de células estaminais hematopoiéticas. No entanto, devido aos efeitos secundários destas terapias, a comunidade científica tem sentido necessidade de desenvolver novas estratégias terapêuticas (Costa, *et al.*, 2015). Vários estudos bioquímicos demonstraram, ao longo dos anos, que a HbF interfere *in vitro* com a HbS, diminuindo a sua polimerização e aumentando a sua solubilidade em condições de desoxigenação (Bauer, Kamran & Orkin 2012; Frenette & Atweh, 2007). Assim, baseado nestas observações, foi proposto que a indução farmacológica da produção de HbF possa ser uma estratégia terapêutica eficaz, de forma a melhorar os sintomas da SCA. Quando, no final dos anos 70, os diferentes genes da globina foram clonados e os mecanismos responsáveis da sua regulamentação foram elucidados durante os anos 80, tornou-se claro que fatores epigenéticos, como metilação do DNA e acetilação das histonas, desempenham papéis importantes na regulação da expressão génica da globina. Assim, foi proposto que os agentes farmacológicos que alteram a configuração epigenética dos genes da γ -globina podem proporcionar uma viabilidade terapêutica de forma a induzir a expressão da Hb F (Frenette & Atweh, 2007). Em 1980 a 5-azacitidina, um agente desmetilante do DNA, foi o primeiro composto utilizado para aumentar a expressão da HbF através do silenciamento epigénico dos genes da γ -globina nos adultos (Bauer, *et al.*, 2012).

O fundamento desta abordagem foi baseado na descoberta de que os transcritos ativos nos genes da β -globina nos adultos são hipometilados, enquanto que os genes não transcritos nos genes da γ -globina fetal são hipermetilados na vida adulta. Em contraste, na vida fetal os genes adultos da β -globina são hipermetilados e os genes da γ -globina são hipometilados (Frenette & Atweh, 2007). Demonstrou-se a capacidade deste composto em estimular a produção de HbF em doentes com SCA e β -talassémia, mas apesar da sua capacidade este composto não foi muito usado em ensaios clínicos devido ao seu poder carcinogénico (Sankaran, 2011). Devido a este facto, desenvolveram-se estudos com outros compostos como a hidroxiureia, um agente quimioterapêutico específico da fase

S, que não inibe a metiltransferase do DNA e que também mostrou aumentar os níveis de HbF em babuínos.

A hidroxiureia é um inibidor da ribonucleótido redutase, que foi usado no tratamento de doenças mieloproliferativas. É um fármaco oral que é relativamente bem tolerado e simples de usar. Após a demonstração da sua capacidade de induzir a produção de HbF em babuínos, a hidroxiureia foi testada num pequeno número de ensaios clínicos em adultos com SCA. Um estudo revelou uma diminuição na frequência de crises dolorosas e síndrome torácica aguda e uma redução nas necessidades de transfusão e hospitalizações em adultos com SCA moderada a grave, após o tratamento com hidroxiureia (Frenette & Atweh, 2007). A hidroxiureia é, neste momento, o único medicamento aprovado pela FAO/OMS (FAO – *Food and Agriculture Organization*; OMS – Organização Mundial de Saúde) para tratamento de doentes com SCA.

Outras drogas têm sido propostas. Assim, em meados dos anos 80 demonstrou-se que crianças filhas de mães diabéticas apresentavam falha no *switch* da hemoglobina fetal para hemoglobina adulta, pelo que foi sugerido que pequenas cadeias de ácidos gordos são capazes de induzir a HbF através da inibição de histonas desacetilases. A inibição destas histonas pode potenciar o aumento de HbF *in vitro*, pelo que estes compostos poderão ter um efeito clínico também *in vivo* (Sankaran, 2011).

Muito recentemente, foi observado que na diferenciação e desenvolvimento dos eritrócitos maduros são expressos vários microRNAs dependendo do estado de diferenciação e estes são fatores de transcrição importantes na produção de HbF. Após tratamento de doentes SCA com hidroxiureia verifica-se o aumento da expressão de alguns microRNAs, os quais são considerados agentes indutores de HbF mediados por hidroxiureia (Figura I.5) (Saki, *et al.*, 2016).

A 2,3-adenosina-dialdeído é um inibidor das DNA metiltransferases e verificou-se que era capaz de induzir a HbF em células K562, em células progenitoras hematopoiéticas humanas, pelo que os autores concluíram que poderia ser usada no tratamento de doentes com SCA e β -talassémia (Costa, *et al.*, 2015).

Outras drogas que têm sido alvo de estudos epigenéticos são os inibidores das histonas desacetilases; o butirato mostrou ser capaz de aumentar a expressão de γ -globina através do aumento da acetilação das histonas e baixando os níveis de metilação do DNA na região do promotor (Costa, *et al.*, 2015).

Também recentemente foi demonstrado que a pomalidomida é capaz de induzir HbF em cultura de células de murganhos com SCA. No entanto, esta droga causa trombose e pancitopenia, pelo que serão necessários mais estudos para se poder usar em humanos (Lowrey, 2016).

I.4.2. Gene *HBG2* (cromossoma 11p)

Em indivíduos não anêmicos do Norte da Europa, o SNP (*Single Nucleotide Polymorphism*) *XmnI*, localizado no cromossoma 11p15, na posição -158, a montante do início da transcrição do gene *HBG2* (*Hemoglobin Subunit Gamma 2*), agora denominado rs7482144, representa 13 a 32% de variação dos níveis de HbF (Cao, Moi & Galanello, 2011; Friedrich, *et al.*, 2016). No entanto, o efeito da alteração -158 C>T não é muito consistente, tendo sido associado também a níveis normais de HbF, tanto em indivíduos normais como em portadores de β -talassémia (Gallienne, *et al.*, 2012). Em contraste, em condições de *stress* eritropoiético, como em homozigóticos para β -talassémia ou em doentes com SCA, esta alteração pode levar a uma elevada produção de HbF, desempenhando assim um papel importante na melhoria do fenótipo associado a estas patologias (Cao, *et al.*, 2011).

A mutação que causa a SCA está em desequilíbrio de *linkage* com cinco haplotipos definidos por um SNP no *cluster* do gene da β -globina. Um destes SNPs inclui o *XmnI*, pelo que estes haplotipos têm sido associados ao aumento dos níveis de HbF e fenótipo clínico menos grave (Steinberg, 2009). Os mecanismos destas associações ainda não estão bem definidos, mas pode estar relacionado com o aumento da expressão do gene $G\gamma$ por impedimento de ligação de proteínas silenciadoras no seu promotor (Cao, *et al.*, 2011).

I.4.3. Gene *BCL11A* (cromossoma 2p)

A identificação de fatores de transcrição importantes no silenciamento dos genes da γ -globina tem sido um alvo importante na especificidade da terapêutica em doentes com hemoglobinopatias. O fator transcricional BCL11A (*B-cell CLL/lymphoma 11A*) é o mais potente repressor da produção de hemoglobina fetal, uma vez que a sua expressão parece estar correlacionada com as diferentes fases do desenvolvimento. Nas células eritróides do fígado fetal expressam-se níveis elevados de γ -globina, ao contrário do BCL11A que

se expressa a níveis muito reduzidos ou até nulos (Sankaran, 2011). Existem estudos que mostram que na ausência deste gene a concentração de HbF aumenta e o efeito na eritropoiese é mínimo (Lettre & Bauer, 2016). O BCL11A tem um papel importante no desenvolvimento dos neurónios, dos linfócitos B e nas células dendríticas e provavelmente também nas células estaminais hematopoiéticas e nos precursores pancreáticos (Lettre & Bauer, 2016).

O mecanismo pelo qual o fator BCL11A silencia a expressão da γ -globina ainda não está totalmente compreendido. No entanto, sabe-se que o BCL11A interage diretamente com a cromatina no *locus* da β -globina nas células primárias eritróides e parece fazer parte de um complexo repressor juntamente com o fator de transcrição GATA-1 e a cromatina NuRD. Quando o BCL11A está ausente, a conformação do *locus* da β -globina altera-se de forma que o LCR ativa a transcrição dos genes da γ -globina (Sankaran, 2011).

Foram descritas variantes genéticas de um nucleótido no gene *BCL11A*. Algumas dessas variantes ou SNPs foram associadas a uma variação positiva da concentração de HbF em doentes com SCA por provocarem a diminuição da expressão do gene *BCL11A*, mantendo cerca de 65% da sua expressão normal. Indivíduos haploinsuficientes para *BCL11A* mostram um aumento da concentração de HbF que varia entre 4,8% e 29,7% (Lettre & Bauer, 2016).

O gene *BCL11A* localiza-se no cromossoma 2p16, e os vários polimorfismos que têm sido estudados em doentes com SCA e outras hemoglobinopatias, incluem o rs11886868, o rs766432, o rs7606173, o rs1427407 e o rs4671393 (Friedrich, *et al.*, 2016; Pereira, *et al.*, 2015). Estes SNPs são variantes intrónicas do gene *BCL11A*.

I.4.4. Gene *HBS1L-MYB* (cromossoma 6p)

Estudos de associação genótipo/fenótipo revelam fortes associações de variantes na região intergénica dos genes *HBS1L* (*HBS1 like translational GTPase*) e *MYB* (*proto-oncogene transcription factor*) com os níveis de HbF. Estes genes localizam-se no cromossoma 6q23 (Thein, Menzel, Lathrop & Garner, 2009). O mecanismo pelo qual estas variantes aumentam os níveis de HbF ainda não se encontra totalmente esclarecido, no entanto, sugeriu-se que isto possa acontecer devido a um efeito na expressão ou na função do *MYB*. A sobreexpressão do *MYB* em células K562 reduz a expressão da γ -globina, e culturas de eritróides primários com elevada expressão de HbF têm baixos

níveis de expressão de *MYB* (Sankaran, 2011). O *MYB* é um fator de transcrição importante na hematopoiese (Bauer, Kamran & Orkin 2012).

Os polimorfismos que têm sido estudados nesta região incluem o rs28384513, o rs4895441, o rs9399137, o rs9389268, o rs9402686 (Friedrich, *et al.*, 2016; Cardoso, *et al.*, 2014). O rs4895441 localiza-se na região a montante do gene, enquanto o rs9399137 e o rs9389268 são variantes intrônicas.

I.4.5. Gene *KLF1* (cromossoma 19p)

Dado que as variantes nos *locus* dos genes *BCL11A*, *HBS1L-MYB* e *HBG2* explicam cerca de 50% da variação nos níveis de HbF, foi levantada a hipótese de que deverão existir mais fatores envolvidos neste mecanismo (Gallienne, *et al.*, 2012). Assim, recentemente, foi identificado outro *locus* importante quando se associaram variantes pontuais no gene *KLF1* em algumas famílias com HPFH (Borg, *et al.*, 2010; Satta, *et al.*, 2011).

O *KLF1* (*Kruppel like factor 1*) é um fator de transcrição eritróide essencial, identificado em 1992 por Miller e seus colaboradores, que se liga a um importante local de ligação ao DNA, o elemento CACCC no promotor do gene da β -globina (Perkins, *et al.*, 2016). Este fator regula o desenvolvimento dos eritrócitos, incluindo a expressão do gene da β -globina, a síntese do heme, o citoesqueleto e a membrana celular, sendo também necessário na ligação à cromatina entre o LCR da β -globina e os promotores (Vinjamur, *et al.*, 2016). Para além disso, estudos recentes demonstraram que o *KLF1* tem um papel importante na regulação do *switch* da expressão da hemoglobina fetal para a hemoglobina adulta através da ativação direta do gene da β -globina e pela repressão indireta da expressão do gene da γ -globina nos progenitores eritróides do adulto através da regulação do *BCL11A* (Siatecka & Bieker, 2011).

O *KLF1*, e os outros modificadores genéticos acima referidos, têm sido alvo de bastantes estudos em indivíduos com hemoglobinopatias devido à sua interação com gene da β -globina e ao efeito que produzem nos níveis de HbF. Há a perspetiva que possam vir a ser usados como alvos terapêuticos na modulação de expressão da hemoglobina fetal, tal como se mostra na Figura I.5.

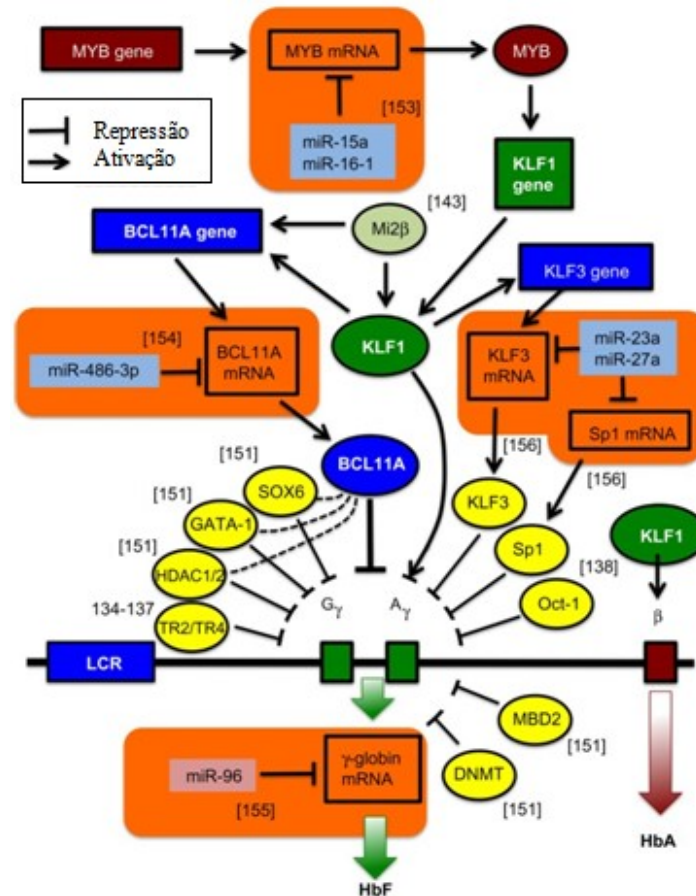


Figura I.5: Novos alvos para a ativação da transcrição do gene da γ -globina. Estão representados os fatores de transcrição que influenciam direta ou indiretamente (positivamente ou negativamente) a transcrição do gene de γ -globina, juntamente com os microRNAs reguladores conhecidos. A elevada expressão de alguns microRNAs pode levar à regulação negativa de repressores de genes de γ -globina (*MYB*, *BCL11A*) e, conseqüente, aumento da produção de HbF (Adaptado de Finotti *et al.*, 2015).

I.5. As vasculopatias cerebrais nas crianças com drepanocitose

O AVC é a principal causa de morbidade e mortalidade nos doentes com SCA, referem-se a um aparecimento súbito focal ou global de deficiência neurológica de origem vascular que dure mais de 24 horas, e podem ser isquêmicos ou hemorrágicos. O AVC surge em cerca de 25% dos doentes com SCA, e em doentes homozigóticos surge com uma prevalência de 10 a 15% antes dos 10 anos de idade, em 11% antes dos 20 anos e em 24% antes dos 45 anos de idade (Switzer, Hess, Nichols & Adams, 2006; Adewoyin, 2015). O risco de AVC aumenta quando os níveis de Hb e de HbF são baixos, quando os glóbulos brancos estão aumentados e quando a pressão sistólica é elevada (Adewoyin, 2015). O AVC hemorrágico ocorre essencialmente entre os 20 e os 29 anos de idade e está associado a níveis baixos de hemoglobina e a leucócitos aumentados. A avaliação do risco de AVC é feita com a determinação da velocidade do fluxo sanguíneo cerebral através do doppler transcraniano e de ultrassonografia (Adewoyin, 2015). Os adultos têm

uma velocidade de fluxo sanguíneo nas artérias cerebrais de aproximadamente 60 m/s, enquanto em crianças sem anemia os valores são de 90 m/s. Na SCA a média do referido fluxo é bastante mais elevada, 130 m/s. O risco de crianças com SCA desenvolverem AVC aumenta com o aumento das velocidades de fluxo sanguíneo (Adams, 2007). Crianças com velocidades de fluxo inferiores a 170 m/s são consideradas normais, entre 170 e 199 m/s apresentam um risco moderado e acima de 200 m/s estão associadas a um maior risco de desenvolver AVC (Adams, 2007).

Os enfartes subclínicos, também chamados enfartes silenciosos, têm uma prevalência que varia entre 17 e 35% em doentes com SCA e destes em cerca de 37% ocorrem antes dos 14 anos de idade (Switzer, *et al.*, 2006). O enfarte silencioso define-se por uma ressonância magnética nuclear alterada, tem como fatores de risco os níveis baixos de hemoglobina e a pressão sistólica elevada, e está associado à disfunção neuropsiquiátrica nestes doentes e ao aumento do risco para desenvolver AVC (Adewoyin, 2015).

O efeito histopatológico mais comum nos AVCs associados a SCA é o dano do endotélio das artérias, induzindo a proliferação e deposição da fibrina e formação do trombo. O espessamento das artérias é devido à proliferação de fibroblastos e células de músculo liso que ocorre como consequência de lesão endotelial recorrente por eritrócitos (Switzer, *et al.*, 2006).

A aderência de células falciformes ao endotélio promove a atividade de fatores de transcrição e vasoconstritores, como a endotelina-1 (ET-1). O relaxamento vascular é inibido e existe um aumento na expressão de moléculas de adesão superficial que promovem ainda mais a interação eritrócito-endotélio. Além disso, a hemoglobina livre inativa o óxido nítrico (NO), que tem ação vaso-dilatadora, aumentando ainda mais o tônus vascular, pelo que os doentes com SCA apresentam um estado pró-coagulante e pró-inflamatório (Figura I.6). O resultado final é a remodelação da parede vascular e a vasculopatia (Switzer, *et al.*, 2006). Consequentemente, o AVC nestas crianças é o resultado da estenose progressiva das artérias, estando a sua extensão relacionada com a gravidade da estenose subjacente (Switzer, *et al.*, 2006).

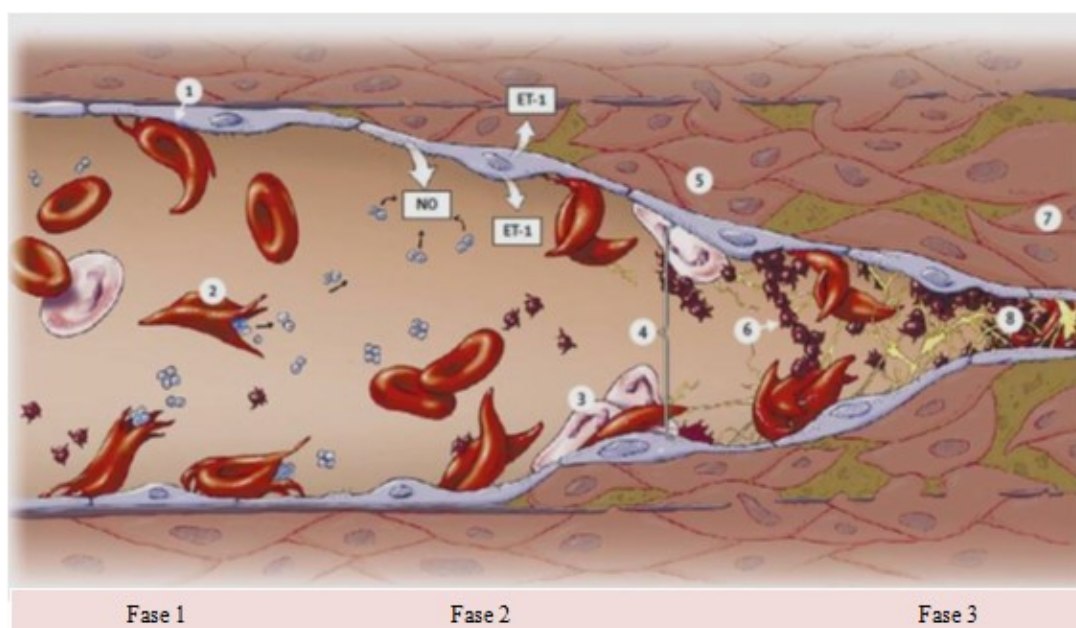


Figura I.6: Vasculopatia e AVC na anemia das células falciformes. Os drepanócitos possuem propriedades diferentes dos eritrócitos normais; têm grande capacidade de adesão às paredes dos vasos sanguíneos (1) e sofrem hemólise (2) precoce. O revestimento celular dos vasos sanguíneos (o endotélio) é ativado e expressa uma série de moléculas específicas que promovem a adesão de outras células sanguíneas, tais como; leucócitos (3) e plaquetas (6); e liberação de moléculas com atividade vasoconstritora (endotelina-1, ET-1). A Hb livre proveniente dos drepanócitos que hemolisaram consome óxido nítrico (NO) (gás com ação vasodilatadora) contribuindo, assim, para aumentar o tônus vascular (4). Simultaneamente, há remodelação do tecido vascular devido à proliferação das células musculares lisas e fibroblastos na camada íntima do endotélio (5), o que conduz ao espessamento da parede do vaso. Por último, ocorre vasculopatia (7) e oclusão do vaso (8) (Adaptado de Switzer, *et al.*, 2006).

CAPITULO II
OBJETIVOS

II – OBJETIVOS

A anemia das células falciformes é uma patologia hereditária que apresenta várias manifestações clínicas de gravidade variável, como anemia hemolítica crónica, crises vaso-oclusivas, infeções de repetição, crises de dor aguda, problemas renais e cardiovasculares. A principal complicação é o desenvolvimento precoce de AVC, o qual surge em cerca de 10 a 15% destes doentes antes dos 10 anos de idade. A heterogeneidade fenotípica da drepanocitose tem sido atribuída à ação modificadora de vários fatores ambientais e genéticos. De entre os fatores genéticos incluem-se variantes em genes globínicos e o nível de hemoglobina fetal, uma vez que esta é inibidora da polimerização de HbS. Os níveis de HbF são também eles, em parte, condicionados por fatores genéticos tais como variantes nos genes *HBG2*, *BCL11A* e *HBSIL-MYB* e, possivelmente, *KLF1*. É sabido que as variantes nestes genes são diferentes entre populações com origem geográfica distinta. Uma vez que os doentes com SCA na nossa população são, maioritariamente, de ascendência africana, levantaram-se as seguintes questões:

- Existe relação entre a presença dos polimorfismos nos referidos genes e os níveis de HbF na nossa população com SCA?
- Existe relação entre o nível de HbF e o grau de gravidade da doença, na nossa população de estudo, nomeadamente com o grau de hemólise e com o risco de desenvolver AVC?
- Existe relação entre a presença dos genótipos referidos e os fenótipos hematológicos, bioquímicos e manifestações clínicas na nossa população de estudo?
- É possível determinar um perfil de marcadores hematológicos, bioquímicos e/ou genéticos de risco para AVC nesta patologia?

Assim, os principais objetivos deste trabalho consistiram em:

- Estudar, a nível hematológico, bioquímico e molecular, um grupo de crianças com SCA que apresentam manifestações clínicas de gravidade variável. Pretendeu-se caracterizar diversas variantes genéticas nos genes *HBG2*, *BCL11A* e *HBSIL-MYB* e estudar a sua possível associação com o nível de hemoglobina fetal, grau de hemólise e risco de AVC.
- Implementar no laboratório uma estratégia rápida e pouco dispendiosa de procura de variantes no gene *KLF1*, através de amplificação por PCR-Longo do referido gene, seguido de análise por sequenciação de nova geração (*Next Generation*

Sequencing, NGS) e validação por sequenciação de Sanger, considerado o método de referência.

- Avaliar qual o papel das variantes encontradas no gene *KLF1* na modulação desta patologia e quais as suas interações com as variantes encontradas nos genes *HBG2*, *BCL11A* e *HBS1L-MYB*.
- Avaliar se a estratégia de análise do gene *KLF1* por NGS será uma mais-valia no complemento do diagnóstico de doentes com drepanocitose.
- Avaliar a possibilidade de definir um perfil de marcadores hematológicos, bioquímicos e/ou genéticos de risco para o desenvolvimento de AVC nesta patologia o que permitiria aplicar uma terapêutica de prevenção personalizada de forma a evitá-los.

CAPITULO III
MATERIAL E MÉTODOS

III - MATERIAL E MÉTODOS

III.1. Amostra

III.1.1. Amostra populacional

A amostra populacional incluída neste estudo foi constituída por um total de 67 crianças, com idade superior a três anos e homozigóticas para a mutação que dá origem à anemia das células falciformes. O genótipo característico nalguns destes indivíduos já tinha sido previamente confirmado no decurso de um projeto anterior de âmbito mais alargado (FCT - PIC/IC/83084/2007) que decorreu no Grupo de Investigação e Desenvolvimento (I&D) em hemoglobinopatias, metabolismo do ferro e patologias associadas da Unidade de Investigação do Departamento de Genética Humana do Instituto Nacional de Saúde Doutor Ricardo Jorge (INSA).

As crianças são de origem africana, sendo os seus progenitores provenientes de várias regiões geográficas, tais como: Angola, Cabo Verde, São Tomé e Príncipe, Guiné-Bissau e Nigéria. Apenas um dos progenitores é de origem portuguesa proveniente do Alentejo. As crianças em estudo foram provenientes da consulta de pediatria de quatro Hospitais situados na região de Lisboa: Hospital A (n=24), Hospital B (n=24), Hospital C (n=18) e Hospital D (n=1).

A amostra foi constituída por 39 crianças do sexo masculino (58,2%) e 28 crianças do sexo feminino (41,8%) e subdividida em três grupos: Grupo Controlo (n=20), Grupo Risco (n=32) e Grupo AVC (n=15). O Grupo AVC incluiu as crianças que já tinham desenvolvido pelo menos um episódio de acidente vascular cerebral. O Grupo Risco incluiu as crianças que apresentam, pelo menos, um dos fatores de risco para o desenvolvimento de AVC, como exame eco-doppler com a velocidade de fluxo sanguíneo no limite superior do normal (170-199 cm/s), ou com a velocidade de fluxo sanguíneo aumentada (>200 cm/s). Também se incluíram neste grupo as crianças que através da ressonância mostraram sinais de terem sofrido enfartes subclínicos ou enfartes silenciosos, ou outra vasculopatia cerebral. No Grupo Controlo incluíram-se as restantes crianças, ou seja, as que apresentavam exame de eco-doppler e ressonância sem alterações e que não tinham sofrido de AVC. Os dados clínicos, parâmetros hematológicos e resultados do eco-doppler e da ressonância foram fornecidos pelos médicos dos respetivos hospitais. Os médicos também obtiveram o consentimento informado devidamente assinado pelos representantes legais das crianças, e quando as amostras de

sangue chegaram ao INSA foram devidamente numeradas de forma a anonimizar a sua identificação.

III.1.2. Amostra biológica

As amostras de sangue foram colhidas nos vários hospitais em tubos com anticoagulante EDTA.K3 e, posteriormente, enviadas para o INSA para extração de DNA. O DNA foi extraído automaticamente através do equipamento de extração de ácidos nucleicos MagNA Pure da Roche e armazenado a 4°C (Sanyo Medicoool Refrigerator).

III.2. Métodos

A identificação molecular da mutação no gene da β -globina (*HBB*), que causa a anemia das células falciformes, foi feita através da técnica de reação em cadeia da polimerase (PCR - *Polymerase Chain Reaction*), seguida de RFLP (*Restriction Fragment Length Polymorphism*). As variantes genéticas estudadas neste projeto foram selecionadas com base em trabalhos desenvolvidos no nosso laboratório (Coelho *et al.*, 2013), e em revisão da literatura através de consulta de bases de dados públicas disponíveis *online*, tais como, NCBI (<https://www.ncbi.nlm.nih.gov/pubmed>) e Ensemble (<http://www.ensembl.org/index.html>). As várias metodologias utilizadas no estudo das diferentes variantes genéticas estudadas encontram-se resumidas na tabela seguinte:

Tabela III.1. – Resumo das metodologias moleculares utilizadas no estudo de cada variante genética

Designação	Gene	Variante genética	Metodologia molecular
Pesquisa da mutação drepanocítica	<i>HBB</i>	HBB:c.20A>T	PCR+RFLP
Variante genética	<i>BCL11A</i>	rs11886868	PCR+RFLP
	<i>HBG2</i>	rs7482144	
	<i>HBSIL-MYB</i>	rs4895441	
	<i>BCL11A</i>	rs4671393	PCR+Sequenciação de Sanger
	<i>MYB</i>	rs9399137	
	<i>MYB</i>	rs9389268	
Estudo do gene <i>KLF1</i>	<i>KLF1</i>	Todo o gene e região promotora	PCR-longo + NGS + Sequenciação de Sanger

III.2.1. Amplificação de DNA genómico pela técnica de *Polymerase Chain Reaction* (PCR)

A reação em cadeia da polimerase (PCR) foi desenvolvida em 1983 por Kary Mullis, e é a técnica de amplificação de ácidos nucleicos mais utilizada no diagnóstico *in vitro* de doenças genéticas.

Esta técnica ocorre em três passos principais que se repetem em ciclos sequenciais. Primeiro o DNA de cadeia dupla é desnaturado em cadeias simples por aquecimento entre 94 e 96°C. De seguida sequências de nucleótidos específicos de cadeia simples (*primers*) ligam-se a estas cadeias de DNA a uma temperatura relativamente baixa (55-64°C) relacionada com a temperatura de *melting* dos *primers*. Por último ocorre a polimerização dos *primers* e síntese de múltiplas cópias da sequência de DNA alvo (Saiki, *et al.*, 1988). Para isso são necessários os quatro dNTPs (dATP, dGTP, dCTP e dTTP), uma enzima DNA polimerase termoestável com temperatura óptima de funcionamento de 72°C e o respetivo tampão. A enzima mais utilizada no PCR é a Taq DNA polimerase, isolada da bactéria termófila *Thermus aquaticus*.

Estes três passos repetem-se um número limitado de vezes, normalmente 30 a 35 ciclos em termocicladores, dando origem a uma produção exponencial do fragmento amplificado.

Para cada variante genética estudada o DNA genómico foi devidamente amplificado num termociclador (T1 ou T Gradient; Biometra) utilizando *primers* específicos e determinadas condições de PCR, as quais se encontram definidas nos anexos (Tabelas A.2; A.4; A.6; A.8; A.9; A.11- A.16). O produto de PCR foi sujeito a uma electroforese em gel de agarose corado com brometo de etídio (EtBr) e visualizado após exposição a luz ultravioleta. Os géis de agarose (Seakem LE Agarose, Lonza) variaram entre 1 e 2 % (p/v) em tampão TBE 1×, com 6% (v/v) de EtBr (10 mg/mL, Sigma). As condições padrão da electroforese foram 70V durante 45 minutos (BIORAD PowerPac300 ou PowerPac Basic) e a visualização do gel por exposição a luz ultravioleta (Uvitec, FireReader).

III.2.2. *Restriction Fragment Length Polymorphism* (RFLP)

As enzimas de restrição são endonucleases que reconhecem sequências específicas palindrómicas de cerca de 6 nucleótidos e que hidrolisam especificamente cada cadeia da molécula de DNA. O ensaio de restrição enzimática é útil na deteção e identificação de

SNPs (*Single Nucleotide Polymorphism*), uma vez que sob condições específicas a enzima hidrolisa apenas as moléculas de DNA que contêm a sequência de nucleótidos por ela reconhecida. A presença de um SNP pode criar ou eliminar um local de restrição de determinada enzima, sendo necessário o conhecimento da sequência e da alteração genética para a escolha da enzima apropriada.

No ensaio de restrição enzimática utilizou-se, geralmente, 5 U de enzima para 10 µL de produto de PCR. A hidrólise foi feita por incubação à temperatura ótima da respectiva enzima, geralmente 37°C, durante um período mínimo de 3 h (Eppendorf Thermomixer Compact). As condições de cada ensaio de restrição enzimática realizado neste estudo encontram-se resumidas em anexo (Tabelas A.3; A.5; A.7 e A.10).

A verificação dos produtos obtidos após hidrólise foi feita através de electroforese em gel de agarose (Seakem LE Agarose, Lonza) que variou entre 1 e 2 % (p/v) em tampão TBE 1×, com 6% (v/v) de EtBr (10 mg/mL, Sigma). As condições padrão da electroforese foram 70V durante 45 minutos (BIORAD PowerPac300 ou PowerPac Basic) e a visualização do gel por exposição a luz ultravioleta (Uvitec, FireRider).

III.2.3. Sequenciação automática pelo método de Sanger

A sequenciação de Sanger ou método de terminação em cadeia foi desenvolvido em 1977 por um bioquímico britânico Fred Sanger e seus colaboradores. Este método serve para a determinação da sequência de nucleótidos de um fragmento de DNA usando como molde uma cadeia de DNA. O método consiste na síntese de cadeias a partir do fragmento de DNA a ser sequenciado, as quais terminam quando é introduzido um didesoxiribonucleótido (ddNTP) marcado com fluorescência. Para cada ddNTP (ddATP, ddTTP, ddCTP, ddGTP) é utilizado um marcador fluorescente de cor diferente. A separação das cadeias marcadas é feita através de electroforese e assim se determina a sequência de nucleótidos do fragmento de DNA original.

A reação de sequenciação automática foi realizada usando o kit de terminação BigDye V 1.1 (Applied Biosystems), de acordo com as condições do fabricante, e utilizando um termociclador Biometra. As condições utilizadas neste estudo encontram-se resumidas nos anexos (Tabela A.17). A electroforese capilar foi efetuada num sequenciador automático 3130 XI (Applied Biosystems) e os resultados analisados através do programa FinchTV v1.4.0 (Geospiza, Inc) ou ChromasPro v2.6.2 (Technelysium Ply Ltd).

III.2.4. PCR-Longo e Sequenciação de Nova Geração (*Next Generation Sequencing*, NGS)

A sequenciação de nova geração (NGS) é uma tecnologia que utiliza a amplificação clonal e a sequenciação por síntese química. O processo identifica as bases de DNA enquanto simultaneamente as incorpora na cadeia do ácido nucleico. Cada base emite um sinal único de fluorescência e é adicionada à cadeia de forma a determinar a ordem da sequência de DNA. A tecnologia de NGS pode ser usada para sequenciar o DNA de qualquer organismo, e é utilizada na sequenciação desde pequenas regiões alvo até genomas inteiros, sendo útil na investigação e diagnóstico de doenças (<https://www.illumina.com/techniques/sequencing/dna-sequencing.html>).

A preparação das bibliotecas de NGS utiliza transposões com adaptadores que se ligam ao DNA alvo, seguida de tagmentação do DNA em fragmentos e adição dos adaptadores e, por último, PCR para adição de índices e *primers* de sequenciação. A sequenciação das bibliotecas por NGS utiliza ligação aleatória dos fragmentos das bibliotecas através dos adaptadores à superfície da *flowcell*, seguido de PCR por ponte e sequenciação por síntese utilizando ddNTPs terminais reversíveis e com marcação fluorescente.

No estudo em causa foi amplificado todo o gene *KLF1* e região promotora perfazendo cerca de 3,2 Kb através da técnica de PCR-Longo (Tabela A.13).

A técnica de NGS foi realizada na Unidade de Tecnologia e Inovação (UTI) do INSA.

A preparação da *workflow* NGS envolveu 5 passos principais:

- purificação dos produtos de PCR através de *beads* paramagnéticas Agentecourt AmpureXP
- quantificação dos produtos de PCR através da plataforma Qubit e utilizando o kit Qubit dsDNA HS
- preparação das bibliotecas através do sistema Nextera XT e utilizando as condições do fabricante
- sequenciação NGS através do aparelho MiSeq e utilizando uma *flowcell* nano de 500 ciclos
- análise de dados da qualidade da corrida através do programa *Sequencing Analysis Viewer* v1.8.46 e análise de dados da qualidade das amostras através dos programas FastQC v0.11.5 e FastqScreen v0.9.3 (despiste de contaminações); e mapeamento das *reads* e análise de variantes através dos programas *MiSeq reporter* v2.6.2 (aposiçãoamento

BWA e análise de variantes *Genome Analysis ToolKit* (GATK)) e *Integrative Genomics Viewer* (IGV) v2.3.86.

III.3. Análise Estatística por SPSS

A base de dados com a informação dos doentes em estudo, incluindo os parâmetros hematológicos e bioquímicos, os dados do eco-doppler e da ressonância e alguns dados moleculares foi previamente construída no Grupo de I&D em hemoglobinopatias, metabolismo do ferro e patologias associadas da Unidade de Investigação do Departamento de Genética Humana do INSA no âmbito de um projeto anterior (FCT - PIC/IC/83084/2007).

O equilíbrio de Hardy-Weinberg (HWE) é um princípio fundamental em estudos genéticos populacionais, que se baseia no facto de as frequências genotípicas AA, AB e BB para um marcador bialélico autossómico se encontrarem nas proporções p^2 , $2pq$ e q^2 , onde p é a frequência do alelo A e $q=1-p$ é a frequência do alelo B (Graffelman J & Weir BS, 2016).

As contagens alélicas foram efetuadas para toda a população ($n=67$) e o equilíbrio de Hardy-Weinberg foi testado para todos os marcadores moleculares através do teste do χ^2 considerando um grau de liberdade e um nível de significância de 0,05. A hipótese nula a ser testada foi H_0 : A população está em equilíbrio de Hardy-Weinberg. Nestas condições o valor do χ^2 tabelado é 3,84, pelo que se aceita H_0 se o valor do χ^2 calculado for inferior a este valor. Para todas as variantes em HWE as diferenças entre as frequências alélicas e genotípicas foram realizadas considerando como grupo comparador o grupo controlo em função do grupo risco e/ou grupo AVC.

A análise estatística foi realizada através do programa IBM SPSS Statistics v23.

Para se poder utilizar testes paramétricos (ex. ANOVA, *t-student*) na análise estatística o parâmetro em estudo tem de obedecer a dois princípios: verificar a normalidade através do teste Kolmogorov-Smirnov e a homogeneidade de variâncias. Se um destes princípios não se verificar procede-se ao estudo através de testes não paramétricos, os quais incluem o teste de Kruskal-Wallis para mais de dois grupos independentes ou o teste de Mann-Whitney quando apenas se tem dois grupos de amostras independentes para comparar.

Para alguns parâmetros hematológicos (HbF, RDW, HGM, Hb, reticulócitos e leucócitos) da população em estudo foi feita uma primeira abordagem através de análise estatística descritiva para determinar médias, desvio padrão e respetivos quartis. Em seguida,

procedeu-se à análise da normalidade e da homogeneidade das variâncias para cada parâmetro para decidir acerca do teste mais adequado e foram traçados vários *Box-Plots* (gráficos de caixa e bigodes) para poder comparar os valores destes parâmetros nos três grupos em estudo. Para cada *Box-Plot* obtido foi formulada uma hipótese considerando a hipótese nula H_0 como a ausência de associação entre a mediana de cada parâmetro em cada grupo, com um nível de significância de 5%.

Para o valor da percentagem de hemoglobina fetal (HbF %) da população total foi testada a normalidade através do teste Kolmogorov-Smirnov com nível de significância estatística de 5% com correção Lilliefors, uma vez que a amostra é superior a 30; e a hipótese nula H_0 utilizada foi a seguinte: determinado parâmetro hematológico segue uma distribuição normal. Se se verificar a normalidade a seguir procede-se ao teste da igualdade de variâncias, se não prossegue-se o estudo com análise de testes não paramétricos. Foram formulados testes de hipóteses considerando como hipótese nula (H_0) a ausência de associação entre as frequências alélicas ou genotípicas entre os grupos em estudo com um nível de significância para rejeitar a hipótese de 5%.

Para as diferentes variantes genéticas em estudo determinou-se o *odds ratio* e o respetivo intervalo de confiança de 95% com o objetivo de verificar associação entre determinado alelo e a doença. Quando o intervalo de confiança de 95% inclui o 1 significa que o risco é igual nas duas populações, se não incluir o 1 pode-se concluir que existe associação estatisticamente significativa.

CAPITULO IV
RESULTADOS E DISCUSSÃO

IV – RESULTADOS E DISCUSSÃO

A amostra populacional incluída neste estudo foi constituída por um total de 67 crianças, com idade superior a três anos e homozigóticas para a mutação que dá origem à anemia das células falciformes.

As crianças são de origem africana, sendo os seus progenitores provenientes de várias regiões geográficas, tais como: Angola (n=43), Cabo Verde (n=7), São Tomé e Príncipe (n=8), Guiné-Bissau (n=8) e Nigéria (n=1). A amostra foi constituída por 39 crianças do sexo masculino (58,2%) e 28 crianças do sexo feminino (41,8%) e subdividida em três grupos: Grupo Controlo (n=20), Grupo Risco (n=32) e Grupo AVC (n=15).

A população em estudo está em equilíbrio de Hardy-Weinberg para todos os SNPs estudados, ao nível de confiança de 95% ($0,64 \leq \chi^2 \leq 0,98$; $p > 0,05$) e para todas as variantes encontradas no gene *KLF1* ($0,0005 < \chi^2 < 0,61$; $p > 0,05$).

IV.1. Estudo dos parâmetros hematológicos e de hemólise na população em estudo

Os parâmetros hematológicos foram determinados em aparelho automático através dos seguintes métodos de medição: citoquímica, impedância, absorvância, citometria de fluxo e fluorometria. Estes parâmetros incluem os valores dos eritrócitos, dos leucócitos, a fórmula leucocitária, os índices calculados, as plaquetas e os reticulócitos.

Os parâmetros de hemólise, incluem os reticulócitos, a LDH e as bilirrubinas. A LDH e as bilirrubinas foram determinadas em aparelho automático de química clínica através de métodos enzimáticos e colorimétricos, respetivamente.

Como foram feitas várias determinações em fase estável para cada doente em estudo, determinou-se a média dos vários parâmetros para cada um deles, sendo este o valor utilizado para posteriores tratamentos estatísticos (Tabela IV.1). Em seguida procedeu-se à determinação da média, mediana, desvio padrão e limites mínimo e máximo de cada parâmetro, para todos os doentes em estudo.

Nos doentes que estavam medicados com hidroxiureia foram usados os parâmetros de HbF antes do início do tratamento.

Neste trabalho só foram consideradas crianças com idade superior a 3 anos, idade a partir da qual os valores de HbF se encontram estabilizados.

A Tabela IV.1 resume os valores determinados, para os vários parâmetros hematológicos e de hemólise, da população em estudo (n=67).

Tabela IV.1: Parâmetros hematológicos e de hemólise da população em estudo

Parâmetro	Média	Desvio Padrão	Mediana	Mínimo	Máximo
HbS (%)	79,5	12,1	80,0	22,0	97,5
HbF (%)	10,1	6,8	9,3	1,5	27,8
Hb total (g/dL)	8,0	0,9	7,9	6,2	10,4
Eritrócitos ($\times 10^{12}/L$)	3,0	0,5	3,0	2,0	4,5
VGM (fL)	82,1	10,0	81,6	63,8	116,1
HGM (pg)	26,0	6,6	26,8	8,1	39,8
Reticulócitos (%)	10,3	4,2	9,9	1,9	23,4
RDW (%)	21,5	3,1	21,1	15,9	32,1
Plaquetas ($\times 10^9/L$)	423,6	142,8	404,1	134,7	965,5
Leucócitos ($\times 10^9$)	12,4	3,0	12,2	5,7	17,7
Neutrófilos ($\times 10^9$)	5,8	1,8	5,5	2,4	10,4
LDH (U/L)	737,9	355,2	626,2	36,0	1673,3
Bilirrubina total (mg/dL)	3,0	1,6	2,6	0,7	8,3

Estes foram os valores utilizados no tratamento estatístico, sendo que a população foi posteriormente dividida de acordo com os sub-fenótipos da doença: grupo controlo, grupo de risco e grupo AVC.

Alguns autores transformam os valores de HbF (%) em expressões matemáticas que inclui a função logarítmica na tentativa de obter uma distribuição normal (Sebastiani *et al*, 2008). Aplicando o teste Kolmogorov-Smirnov à nossa população para os valores de HbF obtém-se $p=0,009$ e, considerando H_0 como a população que segue uma distribuição normal, sendo $p<0,05$ rejeita-se H_0 , ou seja, a população não seguia uma distribuição normal. A nossa população em estudo não apresentava valores de distribuição normal para a HbF, nem mesmo depois de transformada a função, por isso utilizaram-se os valores de HbF em % e utilizaram-se testes não-paramétricos.

IV.2. Pesquisa molecular de mutação drepanocítica

A confirmação da presença da mutação drepanocítica em homozigotia, responsável pela anemia das células falciformes, já tinha sido previamente realizada em 63 dos 67 indivíduos em estudo, pelo que no presente trabalho apenas foram estudados 4 indivíduos

quanto à presença da referida mutação. A pesquisa da mutação que origina a HbS foi feita por PCR-RFLP. Para cada amostra amplificou-se por PCR um fragmento de 390 pb do exão 1 do gene da β -globina (*HBB*), seguido por digestão enzimática com *Bsu36I*. Esta enzima reconhece a sequência palindrômica 5'...CC▼TNAGG...3' (o nucleótido *wild-type* no local da mutação está sublinhado; ▼ significa o local de restrição), dando origem a dois fragmentos de 213 pb e 177 pb. Na presença da mutação A>T, a sequência não é reconhecida e a enzima não hidrolisa. Assim o fragmento de PCR de 390 pb permanece intacto, como se mostra na Figura IV.1.

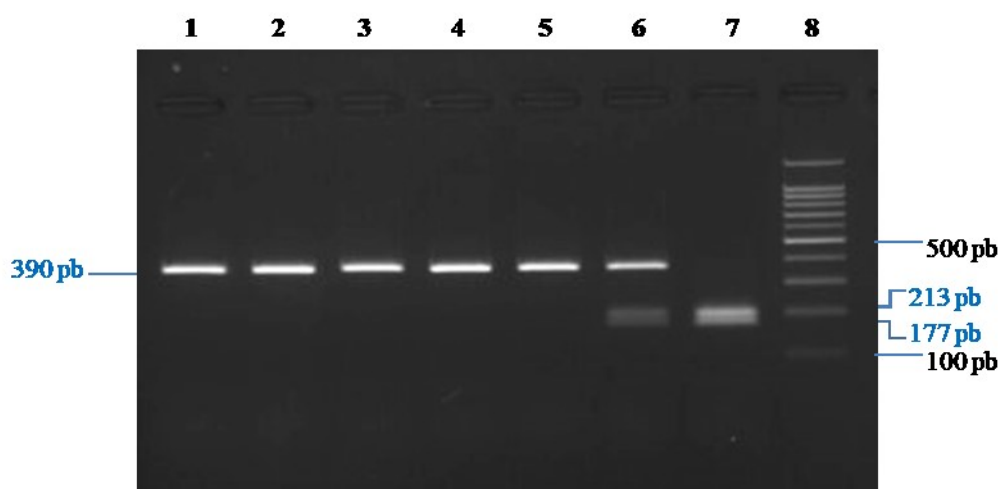


Figura IV.1: Pesquisa de mutação drepanocítica por PCR-RFLP. A imagem mostra a revelação por luz Ultra-Violeta da electroforese em gel de agarose (2% p/v) dos fragmentos de PCR do gene *HBB* após digestão enzimática com *Bsu36I*. A presença da mutação que causa HbS elimina o local de restrição. Nas posições 1 a 4 encontram-se as amostras a serem estudadas. Na posição 5 encontra-se um controle com homozigotia para HbS (Controlo SS). Na posição 6 encontra-se um controle com heterozigotia para HbS (Controlo AS). Na posição 7 encontra-se um controle normal, ou seja, sem a mutação para HbS (Controlo AA). Na posição 8 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). Os números a preto representam os pesos moleculares do marcador de peso molecular e a azul os pesos moleculares dos fragmentos de DNA obtidos após digestão enzimática.

As quatro amostras analisadas confirmaram a mutação em homozigotia para HbS (*HBB*:c.20A>T). Das 67 amostras em estudo apenas uma não confirmou a homozigotia para HbS, sendo que esta confirma ser um composto heterozigótico HbS- β^0 -talassémia (*HBB*: c.118C>T). De realçar que as determinações celulares e moleculares são muito semelhantes no indivíduo HbS- β^0 -talassémia em relação aos encontrados nos indivíduos homozigóticos para HbS, encontrando-se apenas um valor de HbA₂ superior neste indivíduo, típica da β -talassémia, pelo que este indivíduo não foi excluído do estudo.

IV.3. Caracterização molecular de regiões polimórficas reconhecidas como potencialmente modificadoras

IV.3.1. *HBG2*: SNP rs7482144

O SNP rs7482144 (11:5254939G>A) localizado no gene *HBG2* no cromossoma 11 é uma variante da região 5'-UTR e foi estudada em 10 indivíduos por PCR-RFLP, uma vez que os outros 57 indivíduos já se encontravam estudados. O produto de PCR amplificado foi um fragmento de 677 pb. A enzima de restrição utilizada *XmnI* reconhece a sequência palindrômica 5'...GAANN▼NNTTC...3' (o nucleótido variante no SNP está sublinhado; ▼ significa o local de restrição). Na presença do alelo ancestral G, a enzima não reconhece o local de restrição pelo que não hidrolisa o fragmento amplificado, obtendo-se o produto inicial de 677 pb. Na presença do alelo variante A, a enzima reconhece o local de restrição hidrolisando o fragmento de PCR amplificado em dois fragmentos mais pequenos de 462 pb e 215 pb.

Na Figura IV.2 apresenta-se o resultado do fragmento de PCR amplificado e na Figura IV.3 o resultado da digestão enzimática com *XmnI* para determinação do genótipo rs7482144.

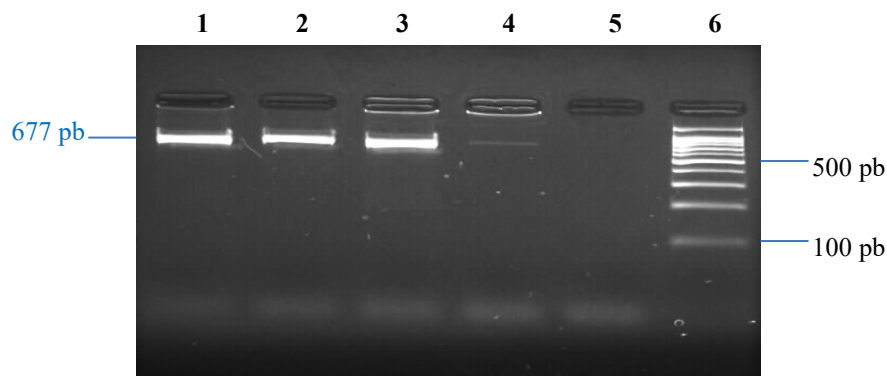


Figura IV.2: Resultado do produto de PCR amplificado do gene *HBG2* para pesquisa do SNP rs7482144. Nas posições 1 a 4 encontram-se algumas das amostras estudadas. Na posição 5 encontra-se o controlo negativo do PCR. Na posição 6 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA amplificado.

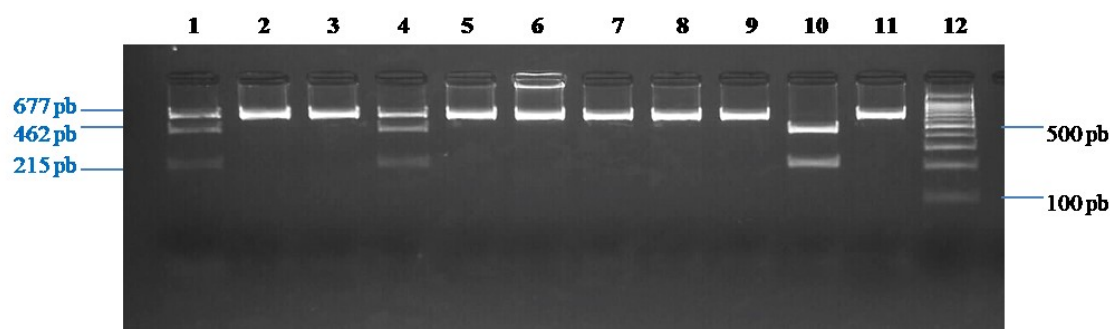


Figura IV.3: Resultado da digestão enzimática do fragmento de PCR amplificado do gene *HBG2* com a enzima de restrição *XmnI* para pesquisa do SNP rs7482144. Nas posições 1 e 4 encontram-se amostras heterozigóticas para o SNP rs7482144 com genótipo CT (677 pb, 462 pb e 215 pb). Nas posições 2, 3, 5, 6, 7, 8, 9 e 11 encontram-se amostras homozigóticas para o alelo ancestral do SNP rs7482144 com genótipo CC (677 pb). Na posição 10 encontra-se uma amostra homozigótica para o alelo variante do SNP rs7482144 com genótipo TT (462 pb e 215 pb). Na posição 12 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA digerido com *XmnI*.

Na Tabela IV.2 resume-se os resultados encontrados para todas as amostras:

Tabela IV.2: Frequências genotípicas e alélicas para o SNP *HBG2* rs7482144

	Genótipo			Alelo	
	CC	CT	TT	C	T
N° Total	50	6	11	106	28
Frequência Genotípica (%) ou Frequência Alélica (%)	74,6	9,0	16,4	79,1	20,9

As frequências alélicas encontradas foram 20,9% para o alelo variante T e 79,1% para o alelo ancestral C. Como este SNP faz parte de um conjunto de SNPs usados para determinar o haplotipo no agrupamento génico da beta-globina e como em África existem descritos 4 haplotipos diferentes: Bantu, Benim, Senegal e Camarões (Miller, *et al.*, 1986), o resultado das frequências encontradas depende das subpopulações estudadas. Na nossa população predomina o haplotipo Bantu (muito frequente na região de Angola) pelo que é concordante uma grande frequência do alelo C e genotípica de CC.

IV.3.2. *BCL11A*: SNPs rs11886868 e rs4671393

O SNP rs11886868 (2:60493111C>T) localizado no gene *BCL11A* (c:386-24278G>A) foi estudado em 4 indivíduos por PCR-RFLP, uma vez que os outros 63 indivíduos já se encontravam estudados. A enzima de restrição utilizada *MboII* reconhece a sequência

palindrômica 5'...GAAGA(N)₈▼...3' (o nucleótido variante no SNP está sublinhado; ▼ significa o local de restrição). Esta sequência ocorre pelo menos uma vez no início do fragmento amplificado, pelo que na presença do alelo ancestral A, a enzima hidrolisa o fragmento de PCR de 659 pb uma vez em dois fragmentos de 433 pb e 226 pb. Na presença do alelo variante G, a enzima hidrolisa o fragmento inicial 2 vezes dando origem a 3 fragmentos mais pequenos de 371 pb, 226 pb e 62 pb. A Figura IV.4 ilustra o PCR do fragmento amplificado do gene *BCL11A* e a Figura IV.5 ilustra o resultado da digestão enzimática com *MboII*.

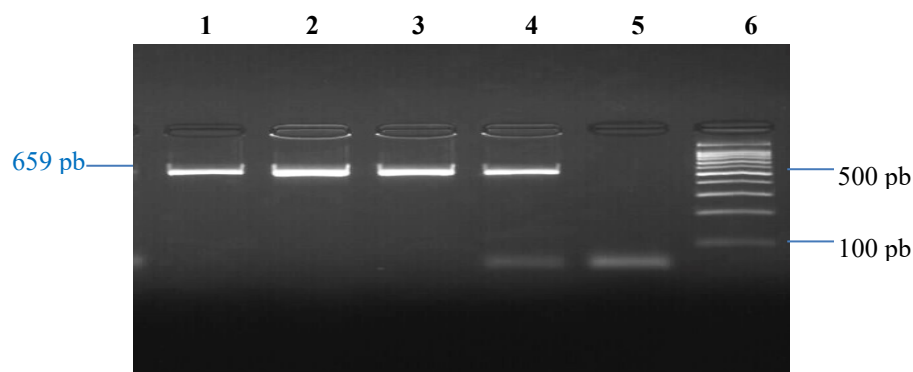


Figura IV.4: Resultado do produto de PCR amplificado do gene *BCL11A* para pesquisa do SNP rs11886868. Nas posições 1 a 4 encontram-se as amostras estudadas. Na posição 5 encontra-se o controlo negativo do PCR. Na posição 6 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA amplificado.

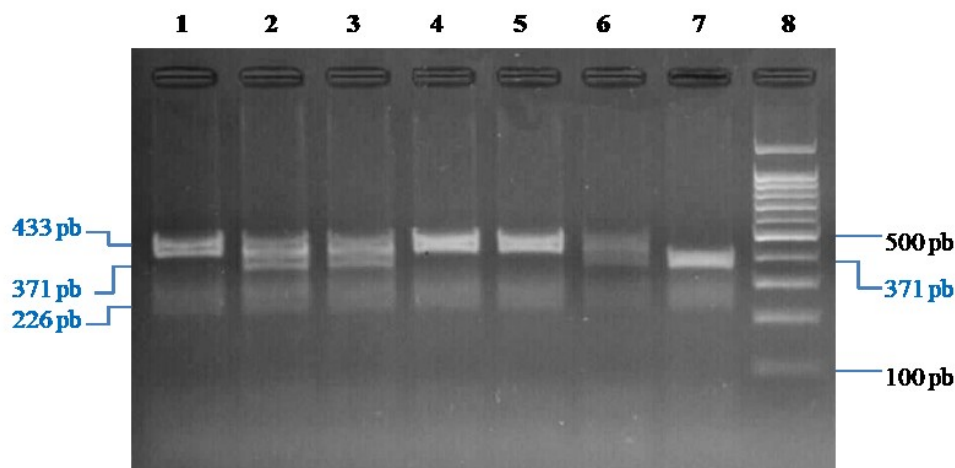


Figura IV.5: Resultado da digestão enzimática do fragmento de PCR amplificado do gene *BCL11A* com a enzima de restrição *MboII* para pesquisa do SNP rs11886868. Nas posições 1 e 4 encontram-se as amostras com genótipo TT (433 pb e 226 pb). Nas posições 2 e 3 encontram-se amostras com genótipo CT (433 pb, 371 pb, 226 pb e 62 pb). Na posição 5 encontra-se uma amostra controlo com genótipo TT. Na posição 6 encontra-se uma amostra controlo com o genótipo CT. Na posição 7 encontra-se uma amostra controlo com o genótipo CC (371 pb, 226 pb e 62 pb). Na posição 8 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA digerido com *Mbo II*.

Nas 4 amostras estudadas para este SNP duas apresentavam o genótipo TT e duas o genótipo CT. No entanto, no tratamento estatístico foram incluídas as 67 amostras. Na Tabela IV.3 resume-se os resultados encontrados para todas as amostras:

Tabela IV.3: Frequências genotípicas e alélicas para o SNP *BCL11A* rs11886868

	Genótipo			Alelo	
	CC	CT	TT	C	T
Nº Total	5	25	37	35	99
Frequência Genotípica (%) ou Frequência Alélica (%)	7,5	37,3	55,2	26,1	73,9
Frequência Alélica (%) na base de dados Ensembl				26	74

A análise do SNP rs4671393 (2:60493816A>G) do gene *BCL11A* localizado no cromossoma 2 foi realizada por amplificação por PCR de um fragmento de 533 pb seguido por reação de sequenciação pelo método de Sanger com o *primer reverse* utilizado no PCR (Tabela A.17). A análise desta variante intrônica foi realizada na população total em estudo, ou seja, nos 67 indivíduos de origem africana com anemia das células falciformes. Na Figura IV.6 apresenta-se o resultado do fragmento de PCR amplificado e na Figura IV.7 o resultado de uma sequenciação para cada genótipo encontrado para o rs4671393.

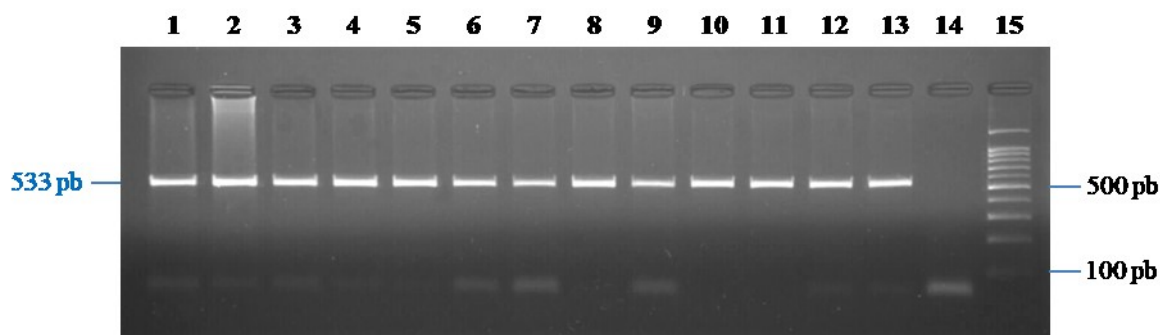


Figura IV.6: Resultado do produto de PCR amplificado do gene *BCL11A* para pesquisa do SNP rs4671393. Nas posições 1 a 13 encontram-se algumas das amostras a serem estudadas. Na posição 14 encontra-se o controle negativo do PCR. Na posição 15 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA amplificado.

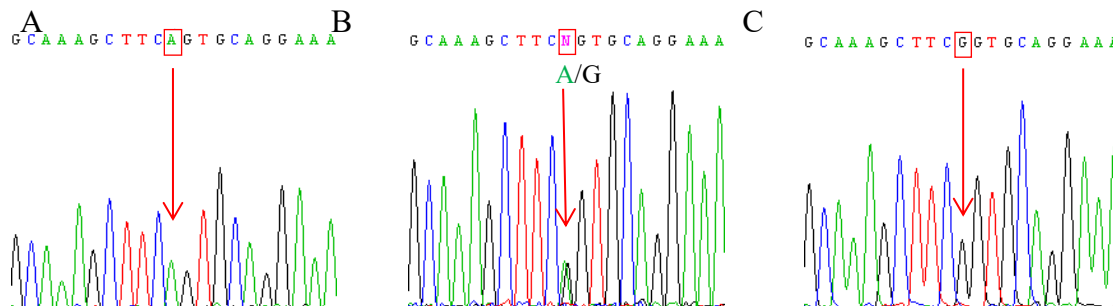


Figura IV.7: Resultado da sequenciação do produto de PCR amplificado do gene *BCL11A* para pesquisa do SNP rs4671393. A - Sequenciação de uma amostra com homozigotia para o alelo A do SNP rs4671393. B - Sequenciação de uma amostra com heterozigotia para o SNP rs4671393 evidenciando a presença dos dois alelos A e G. C - Sequenciação de uma amostra com homozigotia para o alelo G do SNP rs4671393.

Na Tabela IV.4 mostra-se os resultados obtidos na análise do SNP rs4671393:

Tabela IV.4: Frequências genotípicas e alélicas para o SNP *BCL11A* rs4671393

	Genótipo			Alelo	
	GG	AG	AA	G	A
Nº Total	36	24	7	96	38
Frequência Genotípica (%) ou Frequência Alélica (%)	53,7	35,8	10,5	71,6	28,4
Frequência Alélica (%) na base de dados Ensembl				73	27

Como se pode observar através das Tabelas IV.3 e IV.4, as frequências alélicas encontradas para os SNPs rs11886868 e rs4671393, foram semelhantes às descritas na base de dados Ensembl para a população africana em geral.

Este facto é indicativo de não parecer haver uma grande vantagem seletiva positiva pela presença destas variantes na patologia SCA.

IV.3.4. *HBS1L-MYB*: SNPs rs4895441; rs9399137 e rs9389268

O SNP rs4895441 (6:135105435A>G) localiza-se no cromossoma 6 na região intergénica entre os genes *HBS1L* e *MYB*. A análise desta variante foi realizada pela primeira vez na população total em estudo (n=67) por PCR-RFLP com a enzima de restrição *RsaI*. Esta enzima reconhece a sequência palindrómica 5'...GT▼AC...3' (o nucleótido variante no SNP está sublinhado; ▼ significa o local de restrição). No fragmento de PCR amplificado

(778 pb) a enzima reconhece pelo menos um local de restrição, assim na presença do alelo ancestral A, a enzima hidrolisa o fragmento em dois mais pequenos de 594 pb e 184 pb. Na presença do alelo variante G, a enzima reconhece 2 locais de restrição, pelo que após a digestão enzimática se obtêm 3 fragmentos de 494 pb, 184 pb e 100 pb.

Na Figura IV.8 apresenta-se o resultado do fragmento de PCR amplificado e na Figura IV.9 o resultado da digestão enzimática com *RsaI* para determinação do genótipo rs4895441.

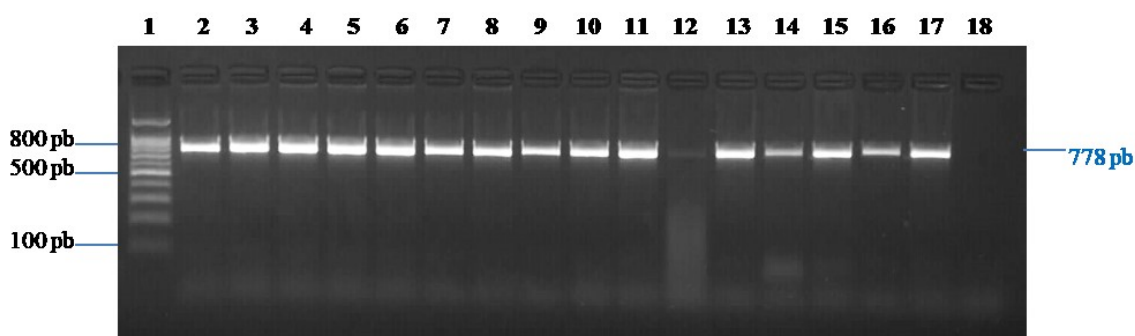


Figura IV.8: Resultado do produto de PCR amplificado da região intergénica dos genes *HBS1L-MYB* para pesquisa do SNP rs4895441. Nas posições 2 a 17 encontram-se algumas das amostras estudadas. Na posição 18 encontra-se o controlo negativo do PCR. Na posição 1 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA amplificado.

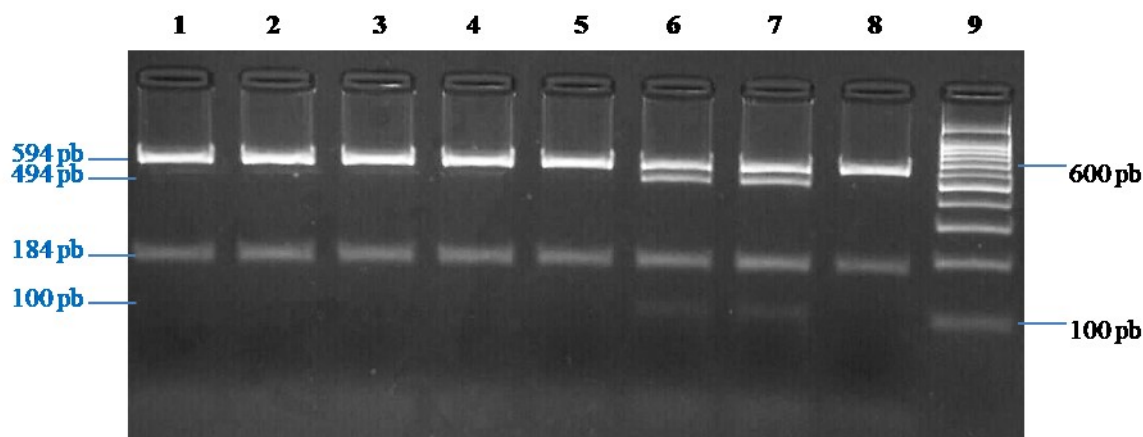


Figura IV.9: Resultado da digestão enzimática do fragmento de PCR amplificado da região intergénica dos genes *HBS1L-MYB* com a enzima de restrição *RsaI* para pesquisa do SNP rs4895441. Nas posições 1, 2, 3, 4, 5 e 8 encontram-se amostras homozigóticas para o alelo ancestral A do SNP rs4895441 com genótipo AA (594 pb e 184 pb). Nas posições 6 e 7 encontram-se amostras heterozigóticas para o SNP rs4895441 com genótipo AG (594 pb, 494 pb, 184 pb e 100 pb). Na posição 9 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA digerido com *RsaI*.

Na Tabela IV.5 mostra-se os resultados obtidos na análise do SNP rs4895441:

Tabela IV.5: Frequências genotípicas e alélicas para o SNP *HBS1L_MYB* rs4895441

	Genótipo			Alelo	
	AA	AG	GG	A	G
Nº Total	58	9	0	125	9
Frequência Genotípica (%) ou Frequência Alélica (%)	86,6	13,4	0	93,3	6,7
Frequência Alélica (%) na base de dados Ensembl				92	8

O SNP rs9399137 (6:135097880T>C) é uma variante intrónica do gene *MYB* localizado no cromossoma 6 e a sua análise foi realizada na população em estudo (n=66), através da amplificação por PCR de um fragmento de 307 pb, seguida de reação de sequenciação pelo método de Sanger com o *primer forward* utilizado no PCR (Tabela A.17).

Na Figura IV.10 apresenta-se o resultado do fragmento de PCR amplificado e na Figura IV.11 o resultado de uma sequenciação para os genótipos encontrados na determinação do SNP rs9399137.

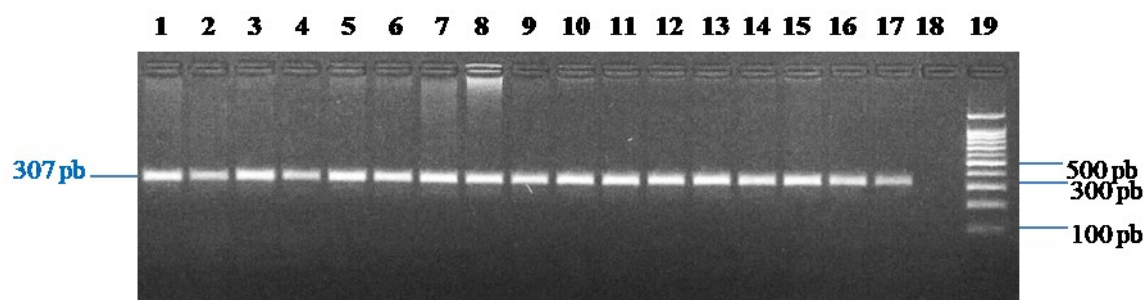


Figura IV.10: Resultado do produto de PCR amplificado do gene *MYB* para pesquisa do SNP rs9399137. Nas posições 1 a 17 encontram-se algumas das amostras estudadas. Na posição 18 encontra-se o controlo negativo do PCR. Na posição 19 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA amplificado.

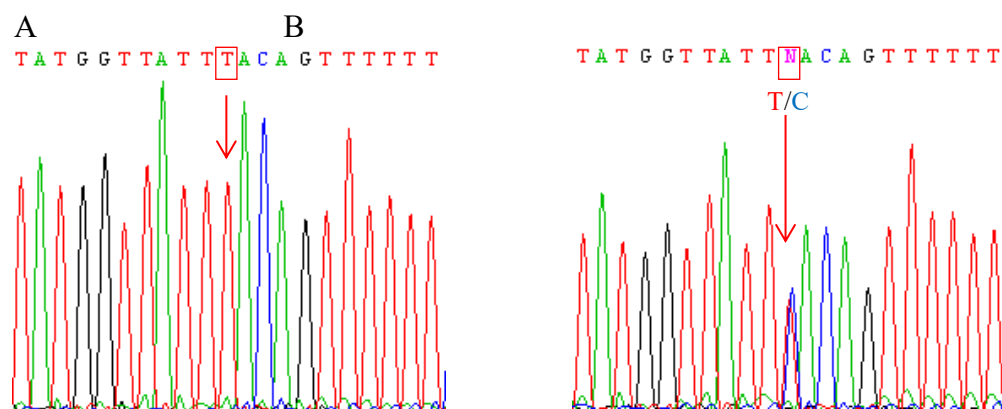


Figura IV.11: Resultado da sequenciação do produto de PCR amplificado do gene *MYB* para pesquisa do SNP rs9399137. A - Sequenciação de uma amostra com homozigotia para o alelo ancestral T do SNP rs9399137. B - Sequenciação de uma amostra com heterozigotia para o SNP rs9399137 evidenciando a presença dos dois alelos T e C.

Na Tabela IV.6 mostra-se os resultados obtidos na análise do SNP rs9399137:

Tabela IV.6: Frequências genotípicas e alélicas para o SNP *MYB* rs9399137

	Genótipo			Alelo	
	TT	TC	CC	T	C
Nº Total	62	4	0	128	4
Frequência Genotípica (%) ou Frequência Alélica (%)	93,9	6,1	0	97	3
Frequência Alélica (%) na base de dados Ensembl				96	4

O SNP rs9389268 (6:135098493A>G) é uma variante intrónica do gene *MYB* e a sua análise foi realizada no total da população em estudo (n=67), através da amplificação por PCR de um fragmento de 344 pb, seguida de reação de sequenciação pelo método de Sanger com o *primer forward* utilizado no PCR (Tabela A.17).

Na Figura IV.12 apresenta-se o resultado do fragmento de PCR amplificado e na Figura IV.13 o resultado de uma sequenciação para cada genótipo encontrado na determinação do SNP rs9389268.

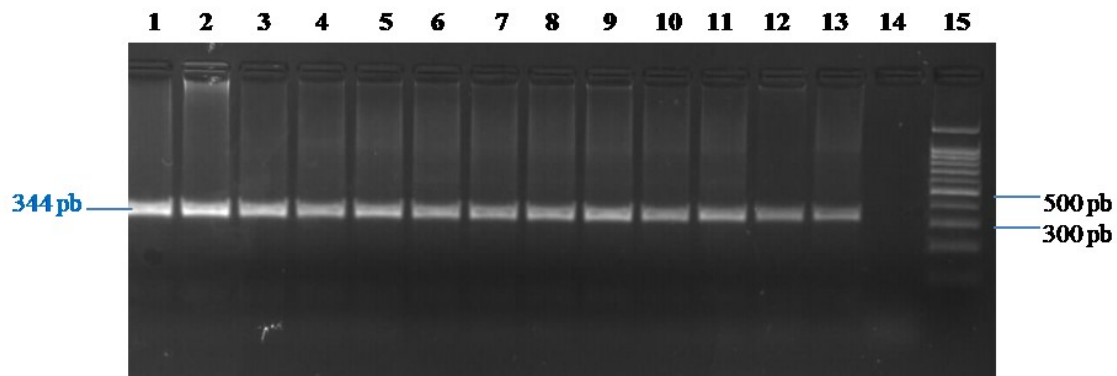


Figura IV.12: Resultado do produto de PCR amplificado do gene *MYB* para pesquisa do SNP rs9389368. Nas posições 1 a 13 encontram-se algumas das amostras estudadas. Na posição 14 encontra-se o controle negativo do PCR. Na posição 15 encontra-se o marcador de peso molecular de 100 pb (100 pb Plus DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA amplificado.

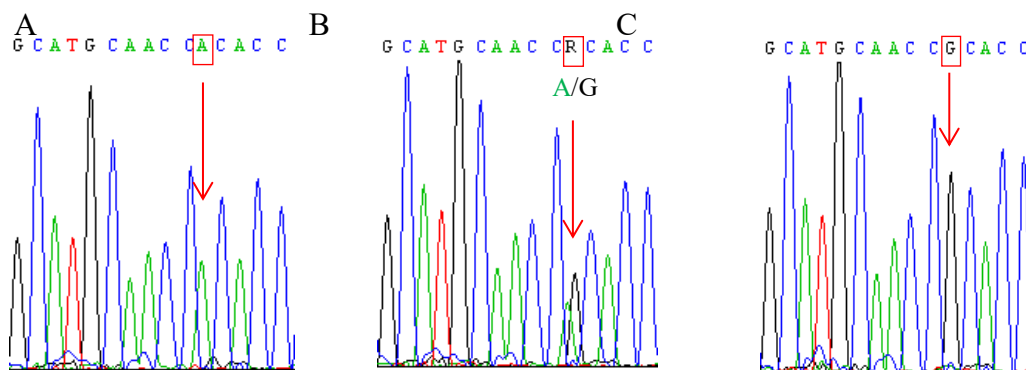


Figura IV.13: Resultado da sequencição do produto de PCR amplificado do gene *MYB* para pesquisa do SNP rs9389268. A - Sequencição de uma amostra com homozigotia para o alelo ancestral A do SNP rs9389268. B - Sequencição de uma amostra com heterozigotia para o SNP rs9389268 evidenciando a presença dos dois alelos A e G. C - Sequencição de uma amostra com homozigotia para o alelo variante G do SNP rs9389268.

Na Tabela IV.7 mostra-se os resultados obtidos na análise do SNP rs9389268:

Tabela IV.7: Frequências genotípicas e alélicas para o SNP *MYB* rs9389268

	Genótipo			Alelo	
	AA	AG	GG	A	G
Nº Total	42	23	2	107	27
Frequência Genotípica (%) ou Frequência Alélica (%)	62,7	34,3	3,0	79,9	20,1
Frequência Alélica (%) na base de dados Ensembl				78	22

As frequências alélicas encontradas para os SNPs rs4895441; rs9399137 e rs9389268, foram semelhantes às descritas na base de dados Ensembl para a população africana em geral (Tabelas IV.5; IV.6 e IV.7 respectivamente).

IV.4. Estudos de associação entre os genótipos e os diferentes sub-fenótipos da doença

IV.4.1. Variantes genéticas versus nível de hemoglobina fetal

Com o objetivo de avaliar a existência de associação entre os vários SNPs estudados e os valores de HbF da população em estudo foram traçados vários *Box-Plots* e aplicado o teste de Mann-Whitney a um nível de confiança de 95%.

Através da análise da Figura IV.14 verifica-se uma tendência para valores mais elevados de HbF no genótipo CC do rs11886868, embora não estatisticamente significativa ($p=0,136$). Também no nosso estudo foi observada uma forte tendência de associação do genótipo AA do rs4671393 com o aumento de HbF, embora não estatisticamente significativo ($p=0,057$), como se mostra na Figura IV.14.

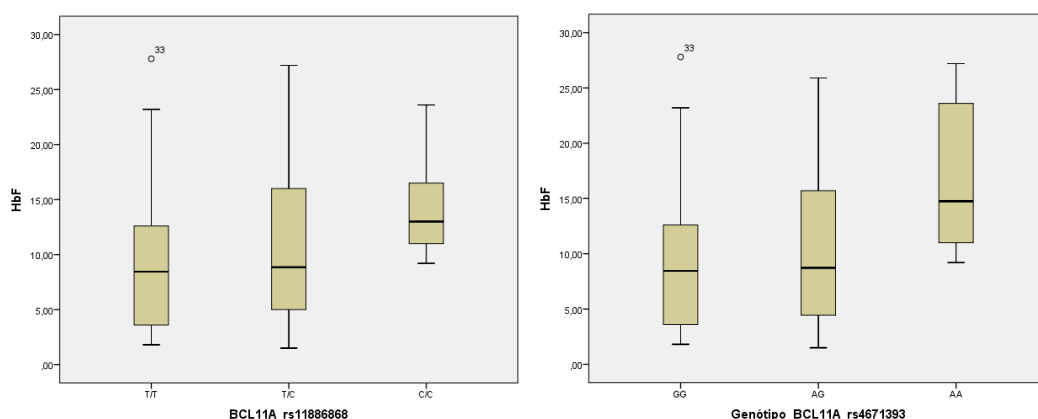


Figura IV.14: Box-Plot dos genótipos do rs11886868 e rs4671393 do gene *BCL11A* em função dos valores de HbF na população em estudo. As caixas representam no limite inferior o 1º quartil e no limite superior o 3º quartil. Fora das caixas está o limite inferior e superior dos valores de HbF de cada grupo. A posição a cheio representa a mediana.

Estes resultados estão de acordo com os descritos por Wonkam *et al.* (2014), numa população com SCA nos Camarões (Wonkam *et al.*, 2014), ou seja, o alelo C do rs11886868 está associado com o aumento dos níveis de HbF. Estes autores também encontraram associação estatisticamente significativa para o rs4671393 no gene *BCL11A*. Também nesta linha de resultados, numa população do Brasil com SCA foi descrita

associação estatisticamente significativa para o rs11886868 e o rs4671393 do gene *BCL11A* e rs9399137 no gene *MYB* com o nível de HbF (Friedrich *et al.*, 2016). Também de acordo, estão os resultados de Cardoso *et al.* (2014), que estudaram vários SNPs numa população brasileira com SCA e encontraram associação com o aumento dos níveis de HbF para o rs4671393 do gene *BCL11A*, rs4895441 do gene *HBSIL-MYB* e para o rs7482144 do gene *HBG2*, embora para este último, uma associação menos forte (Cardoso *et al.*, 2014).

Contrariamente, num estudo semelhante ao nosso, e para uma população semelhante, mas para outros dois SNPs do gene *BCL11A* (rs1427407 e rs7606173), foi reportado a não associação com os níveis de HbF (Pule *et al.*, 2015).

O nosso estudo não revelou associação com os níveis de HbF para o rs4895441 localizado na região intergênica *HBSIL-MYB* ($p=0,181$), nem para outros dois SNPs estudados no gene *MYB*: rs9399137 ($p=0,512$) e rs9389268 ($p=0,182$). Estes resultados levam-nos a concluir que o fator MYB parece ter menos influência nos níveis de HbF na SCA do que o fator BCL11A. A reforçar os nossos resultados encontram-se os estudos recentes de Cardoso *et al.* (2014) e Friedrich *et al.* (2016), que também não encontraram associação estatisticamente significativa entre o SNP rs9399137 e o nível de Hb F.

No entanto, para o rs7482144 do gene *HBG2*, também conhecido por polimorfismo *XmnI*, encontramos uma tendência do genótipo TT a influenciar o aumento dos níveis de HbF ($p=0,096$; Figura IV.15), a qual se torna estatisticamente significativa quando se associam os genótipos que incluem o alelo T ($p=0,031$; Figura IV.15). Os nossos resultados estão de acordo com os descritos na literatura (Gilman & Huisman, 1984; Cardoso, *et al.*, 2014). O aumento dos níveis de HbF, associado ao polimorfismo *XmnI* também foi reportado em doentes com β -talassémia (Pereira, *et al.*, 2015).

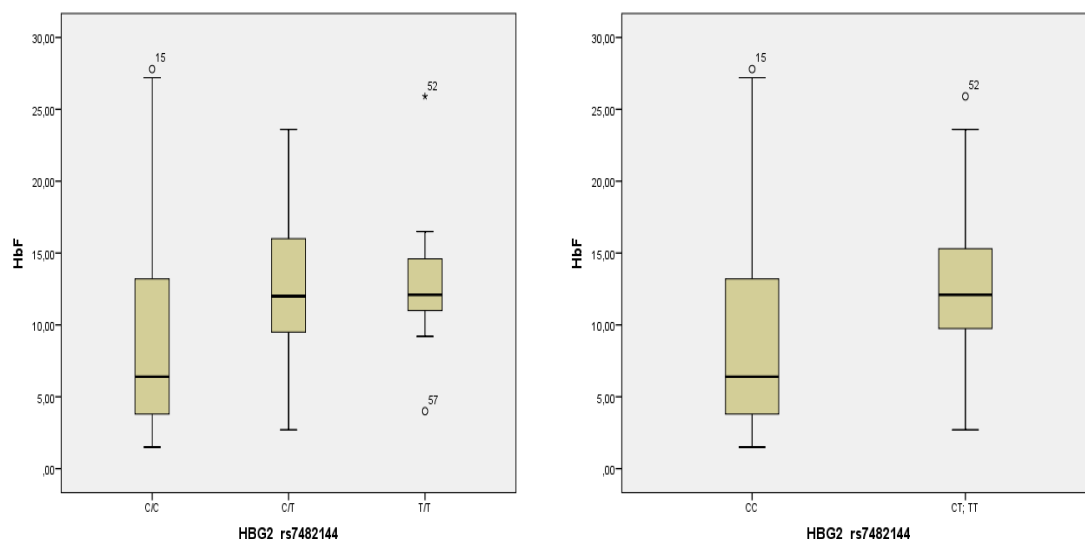


Figura IV.15: Box-Plot dos genótipos do rs7482144 do gene *HBG2* em função dos valores de HbF na população em estudo. As caixas representam no limite inferior o 1º quartil e no limite superior o 3º quartil. Fora das caixas está o limite inferior e superior dos valores de HbF de cada grupo. A posição a cheio representa a mediana. Os pontos com * e ° são os outliers.

Em conclusão, atendendo aos nossos resultados e após uma reflexão sobre o que está descrito na literatura podemos concluir que perante os 3 possíveis genes modificadores do fenótipo da SCA, a saber *HBG2*, *BCL11A* e *MYB* os dois primeiros, *HBG2* e *BCL11A*, parecem exercer uma ação modificadora mais forte sobre o nível de HbF do que o *MYB*. No entanto, a associação entre os diversos SNPs e o aumento da HbF nem sempre é consensual. Dependerá, provavelmente, do número de doentes na amostra, das condições em que são determinados os níveis da HbF (por exemplo: atendendo à idade dos doentes, à toma de hidroxiureia, aos doentes estarem ou não em período de *steady-state*, etc). É de salientar que a grande maioria dos estudos, incluindo o nosso, são realizados usando os parâmetros hematológicos e bioquímicos dos doentes em período de estabilidade entre crises (período em *steady-state*). Pelo contrário, Lettre *et al.* (2008), mostraram associação entre vários SNPs, nomeadamente: o rs11886868, rs4671393 e rs7557939 do gene *BCL11A*; o rs28384513, rs7776054, rs9399137, rs9389268 e rs4895441 dos genes *HBSIL-MYB*; o rs7482144 do gene *HBG2*; e os níveis de HbF, em doentes com SCA quando em crises agudas de dor (Lettre, *et al.*, 2008). Este facto levanta a hipótese de que estes fatores poderão realizar a sua ação de modulação sobre o nível de HbF quando em situação de *stress*.

IV.4.2. Variantes genéticas *versus* risco de AVC

Com o objetivo de avaliar a possível associação de algum dos SNPs estudados ao risco de AVC, foram realizadas várias análises e tratamento estatístico considerando os vários SNPs. Assim, primeiro analisaram-se os genótipos de cada SNP para cada um dos três grupos em estudo (Controlo, Risco e AVC), depois analisou-se a variação dos genótipos do grupo controlo, em função da junção do grupo de risco com o grupo AVC, e do grupo controlo em função do grupo AVC. Por último, foi feita uma análise em função dos alelos presentes em cada SNP. Em relação ao estudo dos alelos foi calculado o intervalo de confiança (IC) a um nível de significância de 95%, de forma a poder determinar o risco relativo.

A Tabela IV.8 resume os valores obtidos para o valor de p através do teste exato de Fisher.

Tabela IV.8: Estudo estatístico para os vários SNPs mostrando as várias combinações dos genótipos e dos alelos com os vários grupos estudados

Gene	Variante	Genótipo (valor p)		Alelos (valor p)		IC (95%)
		Controlo vs AVC	Controlo vs AVC+Risco	Controlo vs AVC	Controlo vs AVC+Risco	
<i>BCL11A</i>	rs11886868	0,248	0,387	0,935	0,534	0,318 – 1,810
	rs4671393	0,237	0,199	0,642	0,574	0,339 – 1,822
<i>HBG2</i>	rs7482144	0,868	0,966	0,511	0,868	0,431 – 2,710
<i>HBS1L-MYB</i>	rs4895441	0,702	0,806	0,712	0,813	0,200 – 3,543
<i>MYB</i>	rs9399137	1,000	0,863	1,000	0,865	0,083 – 8,138
	rs9389268	0,417	0,562	0,334	0,978	0,402 – 2,553

Através da análise da Tabela IV.8 pode verificar-se que todos os valores de p obtidos foram superiores a 0,05 pelo que não se rejeita a hipótese nula, ou seja, não há diferenças estatisticamente significativas entre os grupos estudados, quer quando se avalia os diferentes genótipos, quer a nível dos alelos em estudo, concluindo-se que não se associa, isoladamente, nenhum dos SNPs estudados com o risco de desenvolver AVC. No entanto, verifica-se uma tendência para valores de p mais baixos quando se compara o grupo controlo apenas com o grupo AVC, mas para se poder obter uma associação estatisticamente significativa seria necessário aumentar o número de doentes na amostra. Obtém-se a mesma conclusão quando se avalia o risco relativo, pois como o 1 está

presente em todos os IC, então não é estatisticamente significativo e não há um risco associado ao AVC.

IV.4.3. Estudos de associação entre o nível de Hb e o risco de AVC

Através dos níveis de Hb dos três grupos em estudo traçou-se o gráfico representado na Figura IV.16.

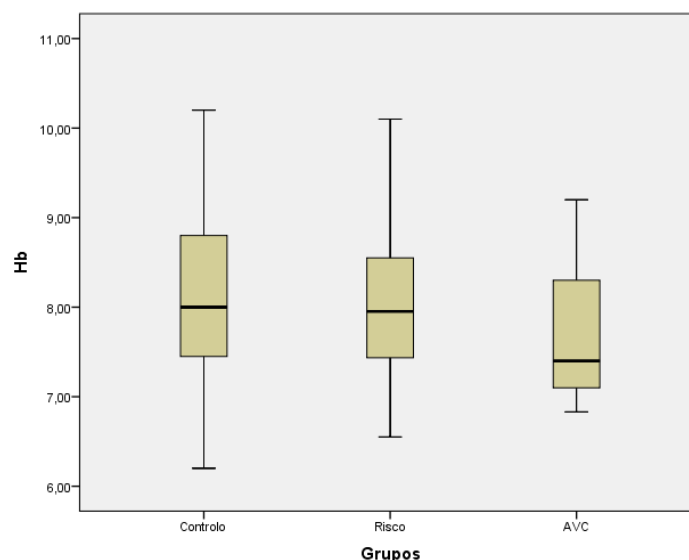


Figura IV.16: Box-Plot dos níveis de Hb (g/dL) para os três grupos em estudo, o grupo controlo, o grupo de risco e o grupo AVC. As caixas representam no limite inferior o 1º quartil e no limite superior o 3º quartil. Fora das caixas está o limite inferior e superior dos valores de HbF de cada grupo. A posição a cheio representa a mediana.

Através da análise da Figura IV.16 observa-se que a população em estudo apresenta em todos os grupos valores de hemoglobina abaixo dos valores de referência para uma população de crianças caucasianas considerada normal (11,0-14,0 g/dL; para crianças entre os 6 meses e os 11 anos de idade), segundo a Norma nº 063/2011 da Direção -Geral da Saúde. Estes resultados sugestivos de anemia estão de acordo com o tipo de população estudada uma vez que todos os grupos têm Anemia das Células Falciformes que é uma anemia crónica. Contudo, os valores de Hb para o grupo AVC revelam um valor de Hb ligeiramente inferior aos outros grupos estudados, o que parece indicar que um grau de anemia mais elevado (ou seja, níveis inferiores de Hb total) se associa tendencialmente a risco para AVC. Este facto também se encontra descrito na literatura (Steinberg, *et al.*, 2003; Elmariah, *et al.*, 2014) e os resultados aqui apresentados reforçam-no. Os níveis baixos de hemoglobina também foram associados ao desenvolvimento de problemas de

baço em doentes com SCA, enquanto história de asma e glóbulos brancos elevados, aumenta o risco de síndrome de dor aguda nestes doentes (Alsultan, *et al.*, 2012).

Estudos semelhantes com os parâmetros de hemólise (LDH, reticulócitos e bilirrubina) não mostraram alterações significativas entre os grupos ($p=0,340$; $p=0,374$ e $p=0,335$, respetivamente) sugerindo que o grau de hemólise não parece influenciar o desenvolvimento de AVC.

No entanto, está descrito na literatura, associação dos marcadores de hemólise (LDH, reticulócitos e bilirrubinas) com a suscetibilidade de desenvolver AVC, numa população Brasileira com SCA (Domingos, *et al.*, 2014).

IV.4.4. Estudos de associação entre o nível de Hb F e o risco de AVC

Através dos valores de HbF dos três grupos em estudo (Controlo, Risco e AVC) traçou-se o gráfico representado na Figura IV.17.

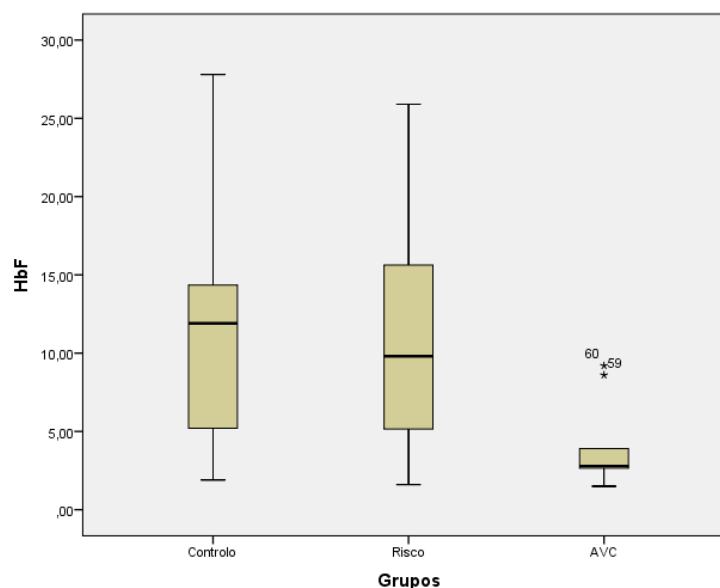


Figura IV.17: Box-Plot dos valores de HbF (%) para os três grupos em estudo, o grupo controlo, o grupo de risco e o grupo AVC. As caixas representam no limite inferior o 1º quartil e no limite superior o 3º quartil. Fora das caixas está o limite inferior e superior dos valores de HbF de cada grupo. A posição a cheio representa a mediana. Os pontos com * são os *outliers*.

Através da análise da Figura IV.17 observa-se, nitidamente, que os valores de HbF para o grupo AVC são muito mais inferiores do que os outros grupos. Aplicando o teste de Kruskal-Wallis de amostras independentes e considerando como H_0 : a distribuição de HbF é a mesma entre os grupos, obtém-se um $p=0,005$, pelo que se rejeita a hipótese nula, e pode dizer-se que os valores de HbF apresentam uma diferença estatisticamente

significativa entre os grupos, e que o grupo AVC tem valores estatisticamente inferiores quando comparado com o grupo controlo. Assim pode concluir-se que o risco de AVC está associado a valores inferiores de HbF. Este conceito de que a baixa HbF é um fator de risco para AVC na SCA já tinha sido proposto por alguns autores (Platt, *et al.*, 1994) e é aqui confirmado. No entanto, outros autores não conseguiram associar os níveis de HbF com as complicações vaso-oclusivas da SCA, mas associaram os valores elevados de HbF com a esplenomegalia persistente em doentes com SCA (Alsultan, *et al.*, 2012). Por outro lado, os valores de HbF elevados também têm sido associados à diminuição dos sintomas em doentes com β -talassémia (Weatherall, 2001; Galanello, *et al.*, 2009). Estes fatores genéticos conhecidos como potencialmente modificadores, juntamente com fatores ambientais e hematológicos, poderão modular o fenótipo dos doentes com SCA.

IV.5. Pesquisa e caracterização de variantes no gene *KLF1*

IV.5.1. Amplificação do gene *KLF1* através da técnica de PCR-longo

O *KLF1* é constituído por dois domínios: o domínio de transativação N-terminal rico em prolina e o domínio C-terminal de ligação ao DNA que reconhece uma região *consensus* 5'CCM CRC CCN (onde R representa A ou G e M representa A ou C). Na região C-terminal existem 3 domínios *Zinc Finger* (Siatecka & Bieker, 2011).

O gene *KLF1*, localizado no braço curto do cromossoma 19 (12,884,423-12,887,181) foi totalmente amplificado com sucesso em 58 amostras, através da técnica de PCR-longo, incluindo as regiões 5'-UTR e 3'-UTR (num total de 3,2 Kb), para posterior análise de variantes através de sequenciação por NGS.

Na amplificação de todo o gene *KLF1* e região promotora utilizou-se o FailSafe PCR System (Epicentre). O FailSafe PCR PreMix utilizado foi o tampão D, cuja mistura inclui vários reagentes desenhados para aumentar a produtividade e especificidade do PCR (tampão, $MgCl_2$, os 4 dNTPs e betaina), sendo utilizada na amplificação de regiões ricas em GC e em fragmentos longos (até aproximadamente 20Kb). A FailSafe PCR Enzyme Mix é uma mistura enzimática que contém um sistema de *proofreading* 3'→5' para aumentar a fidelidade do PCR. Em seguida adicionou-se os *primers* específicos do gene em causa e água e definiram-se as condições de PCR (Tabela A.13).

A Figura IV.18 mostra a amplificação do gene *KLF1*.

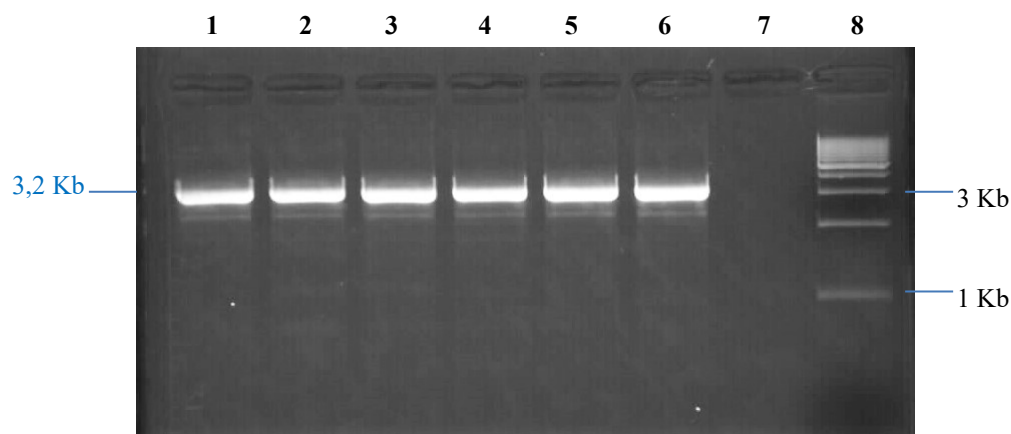


Figura IV.18: Resultado do produto de PCR amplificado do gene *KLF1* para posterior sequenciação por NGS. Nas posições 1 a 6 encontram-se algumas das amostras estudadas. Na posição 7 encontra-se o controlo negativo do PCR. Na posição 8 encontra-se o marcador de peso molecular de 1 Kb (1 Kb DNA Ladder, Bioron). A preto representam-se os pesos moleculares do marcador de peso molecular e a azul o peso molecular do fragmento de DNA amplificado.

IV.5.2. Análise bioinformática dos resultados de NGS e validação dos resultados

O mapeamento das *reads* e análise de variantes obtidas por NGS através dos programas *MiSeq reporter*, *Integrative Genomics Viewer* (IGV), BWA, GATK permitiu encontrar 11 variantes diferentes, das quais uma delas não se encontra descrita na literatura. Na Tabela IV.9 apresenta o resumo das variantes do gene *KLF1*, encontradas por NGS, nas 58 amostras sequenciadas.

Tabela IV.9: Variantes identificadas no gene *KLF1*, através de NGS

Variante	Pos ch38	Alteração Nucleótido	Localização no gene	Tipo variante	Alteração aminoácido
Alteração G>A	12887601	c.-461C>T	5' UTR		
rs3817621 G>C	12887391	c.-251C>G	5' UTR		
rs10407416C>G	12886919	c.87+135G>C	Intrão 1	Intrónica	
rs112631212T>G	12886115	c.115A>C	Exão 2	Missense	p.Met39Leu
rs2072597A>G	12885926	c.304T>C	Exão 2	Missense	p.Ser102Pro
rs2072596A>G	12885686	c.544T>C	Exão 2	Missense	p.Phe182Leu
rs115339342C>T	12885282	c.913+35G>A	Intrão 2	Intrónica	
rs146658904C>G	12884948	c.1026G>C	Exão 3	Missense	p.Gln342His
rs185829666G>A	12884775	c.*110C>T	3'UTR	Conservada	
rs16978757G>A	12884608	c.*277C>T	3'UTR	Conservada	
rs16978754C>T	12884589	c.*296G>A	3'UTR	Conservada	

Como era a primeira vez que este gene estava a ser estudado pelo método de NGS e de forma a poder validar o método, as variantes identificadas foram, posteriormente, confirmadas por PCR-sequenciação de Sanger. As respetivas condições de amplificação e de sequenciação encontram-se em anexo (Tabela A.14; A.15; A.16 e A.17). De seguida mostram-se os resultados da sequenciação por Sanger que permitiu validar os resultados do NGS (Figura IV.19).

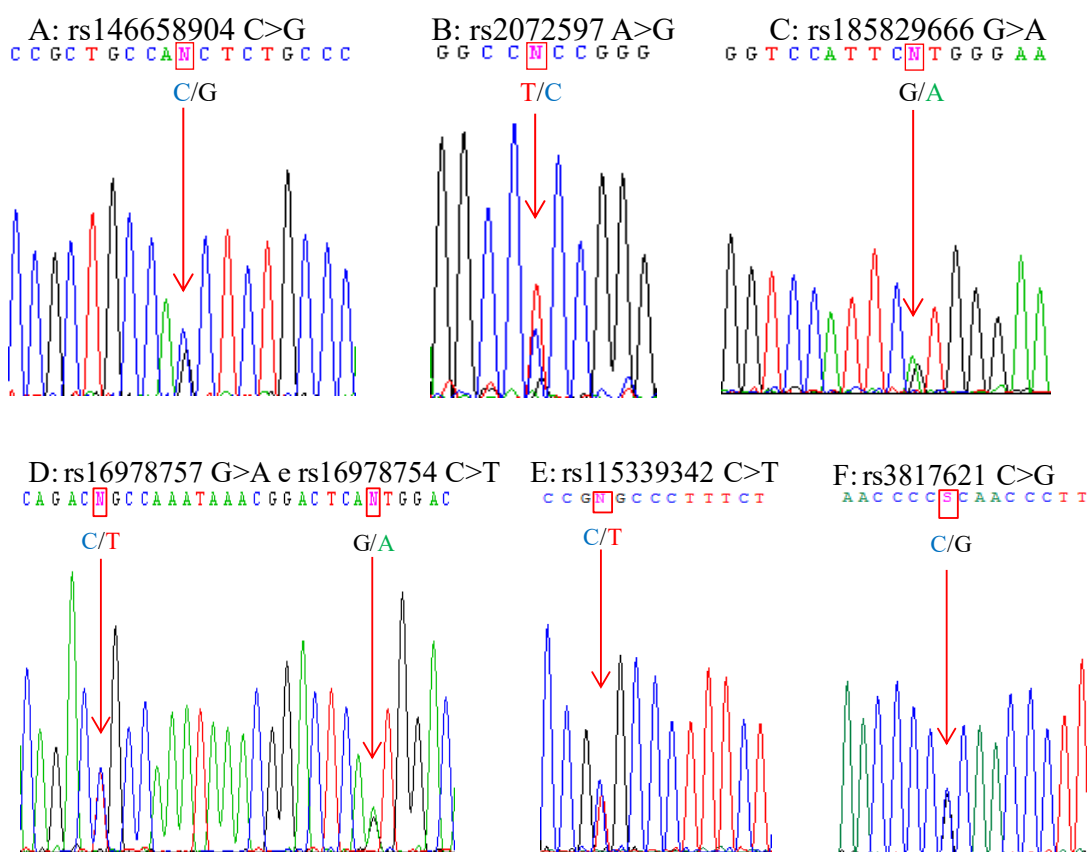


Figura IV.19: Resultado da sequenciação de Sanger que permitiu confirmar as variantes encontradas por NGS na população em estudo, no gene *KLF1*. **A** - Amostra heterozigótica para a variante não descrita rs146658904 C>G (Q342H) mostrando a presença dos dois alelos C e G. **B** - Amostra heterozigótica para a variante rs2072597 A>G mostrando a presença dos dois alelos T e C (sequência *forward*). **C** - Amostra heterozigótica para a variante rs185829666 G>A mostrando a presença dos dois alelos G e A. **D** - Amostra heterozigótica para a variante rs16978757 G>A mostrando a presença dos dois alelos C e T e heterozigótica para a variante rs16978754 C>T mostrando a presença dos dois alelos G e A. **E** - Amostra heterozigótica para a variante rs115339342 C>T mostrando a presença dos dois alelos C e T. **F** - Amostra heterozigótica para a variante rs3817621 C>G mostrando a presença dos dois alelos C e G.

Na Figura IV.20 encontra-se uma representação esquemática do gene *KLF1* com a localização das variantes encontradas.

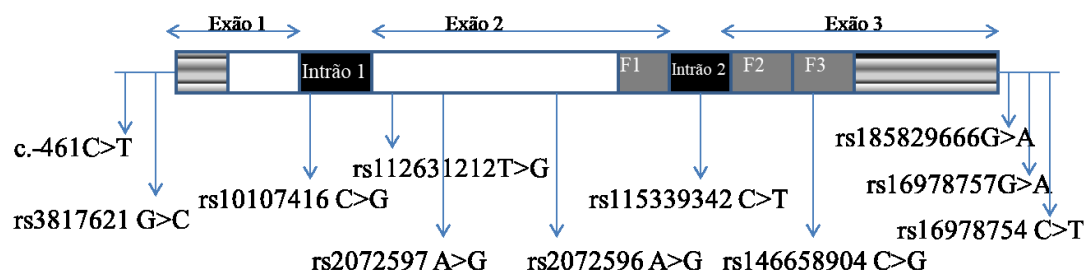


Figura IV.20: Representação esquemática das mutações encontradas na população em estudo no gene *KLF1*. Na figura estão representados os exões e os intrões do gene *KLF1* e a localização das variantes genéticas encontradas. F1, F2 e F3 representam os 3 domínios *Zinc Finger*.

A variante rs10407416C>G (c.87+135G>C) localizada no intrão 1 do gene *KLF1* não foi possível confirmar por PCR-sequenciação de Sanger uma vez que, para o fazer, seria necessário desenhar novos *primers* e otimizar mais um fragmento de PCR.

IV.5.3. Vantagens e desvantagens da utilização da metodologia de NGS

Quando se compara o método NGS com os métodos de sequenciação tradicionais pode apontar-se algumas vantagens e desvantagens em ambos. O NGS é uma técnica mais rápida, menos dispendiosa, dá uma informação em grande escala e podem processar-se muitas amostras simultaneamente. No entanto, quando se trata de sequenciar genomas, ou mesmo exomas, o estudo torna-se bastante moroso e de difícil interpretação e poderá trazer problemas de natureza ética por serem obtidos resultados acidentais cuja interpretação poderá ser ou não fiável. Por isso, sempre que possível, opta-se pelo estudo de um gene ou uma combinação de genes (painel de genes), possivelmente relacionados com a patologia, para se chegar ao diagnóstico clínico.

Como boa prática laboratorial é consensual que cada variante detetada por NGS seja confirmada/validada por sequenciação de Sanger. No nosso caso, os resultados de uma das amostras, apesar de ter sido excluída no método de NGS devido ao reduzido número de *reads*, revelou a presença de 3 variantes, as quais se confirmaram por Sanger. No entanto, nesta amostra a sequenciação de Sanger revelou a presença de mais uma variante

que não foi detetado no NGS. Assim concluiu-se que haverá que repetir a análise por NGS nos casos em que o número de *reads* seja demasiado baixo, pois estas *reads* podem comprometer o número de variantes detetadas. A outra amostra, cujos resultados foram excluídos no controlo de qualidade do NGS devido à baixa qualidade das *reads*, não revelou a presença de variantes no gene *KLF1* através da sequenciação por Sanger. No NGS existiram 6 amostras que revelaram distorção da frequência alélica para todas as variantes identificadas em heterozigotia; estes resultados poderão dever-se a amplificação alélica diferencial ou a eventuais contaminações entre amostras do mesmo amplicão. Estas variantes em heterozigotia foram apenas confirmadas em 5 destes doentes, por sequenciação de Sanger.

Os resultados obtidos por NGS na população em estudo foram bastante concordantes com os confirmados por sequenciação de Sanger, o que confirma que esta técnica tem elevada precisão e exatidão.

IV.5.4. Tipo de variantes encontradas, consequências funcionais e frequências alélicas

Na Tabela IV.10 apresenta-se a frequência alélica determinada para cada uma das variantes identificadas por NGS no gene *KLF1*:

Tabela IV.10: Frequência alélica das várias variantes identificadas no gene *KLF1* através de NGS

Variante	Nº de amostras heterozigóticas	Nº de amostras homozigóticas	Frequência alélica (%) na população em estudo	Frequência alélica (%) na população africana (*)
12887601 G>A	1	0	1,0	ND
rs3817621 G>C	18	5	29,2	19
rs10407416C>G	2	0	2,1	5
rs112631212T>G	6	0	6,3	4
rs2072597A>G	25	14	55,2	44
rs2072596A>G	3	2	4,2	9
rs115339342C>T	1	0	1,0	2
rs146658904C>G	1	0	1,0	0
rs185829666G>A	1	0	1,0	0
rs16978757G>A	13	3	19,8	20
rs16978754C>T	13	3	19,8	20

(*) in <http://www.ensembl.org> ND – não definida

Na grande maioria dos casos as frequências encontradas para as diferentes variantes foram semelhantes às descritas para a população africana em geral. De notar que, estamos a obter a frequência destas variantes para um grupo populacional africano com a doença drepanocitose, enquanto o que é reportado nas bases de dados públicas refere-se à população africana em geral. No entanto, para duas das variantes identificadas (rs3817621G>C e rs2072597A>G), a frequência alélica determinada foi superior à descrita, respetivamente, 29,2% *versus* 19% e 55,2% *versus* 44%. Estes valores, a serem confirmados, podem sugerir que estas duas variantes são benéficas para esta patologia e, portanto, ao longo dos anos tenham sofrido uma seleção positiva.

As variantes rs16978757G>A e rs16978754C>T, ambas localizadas na 3' UTR do gene, parecem estar em *linkage* e, portanto, fazerem parte de algum haplotipo, uma vez que estavam presentes nos mesmos indivíduos, quer em homozigotia, quer em heterozigotia. Mutações no gene *KLF1* têm sido associadas a aumento da HbF e HbA₂, persistência hereditária de hemoglobina fetal e anemia desferitropoiética congénita. Consoante o local da mutação no gene, assim poderá ser atribuída a patologias diferentes. Arnaud *et al.* (2010) identificou duas mutações em heterozigotia no gene *KLF1* num doente com anemia desferitropoiética congénita, uma alteração T>C no exão 2 (c.304T>C; p.Ser102Pro) e uma alteração G>A no exão 3 (c.973G>A; Glu325Lys). A primeira já estava previamente descrita como rs2072597, e como foi também identificada na mãe do doente os autores concluíram que não era patogénica. Esta variante rs2072597 foi encontrada na população em estudo neste trabalho com uma frequência alélica considerada elevada, de 55,2%. A outra mutação que os autores identificaram ainda não se encontrava descrita e através de estudos de expressão celular demonstraram que era a causa da patologia do doente.

A variante rs146658904C>G (c.1026G>C) não foi previamente descrita na literatura, tendo sido identificada por nós em heterozigotia, em apenas um indivíduo. Esta variante é uma mutação *missense*, localiza-se no exão 3 na região do *zinc finger F3*, altera o codão 342 de CAG>CAC e, conseqüentemente, está envolvida na substituição aminoácídica de glutamina por histidina. A glutamina é um aminoácido condicionalmente essencial, de carácter polar neutro, enquanto a histidina é um aminoácido essencial, de carácter básico com carga positiva. As diferentes propriedades bioquímicas destes aminoácidos sugerem que a variante poderá perturbar a estrutura da proteína.

O *KLF1* está associado aos níveis elevados de HbF nos adultos e regula a expressão do gene da hemoglobina fetal através da ativação da expressão do *BCL11A* (Arnaud *et al.*, 2010).

Numa família com HPFH foi identificada uma mutação *nonsense* no gene *KLF1*, p.K288X, a qual elimina os 3 domínios *zinc finger*, de ligação ao DNA, deste fator regulador transcricional eritróide (Borg *et al.*, 2010). Estes autores também encontraram a mutação p.M39L (rs112631212), mas concluíram que é uma substituição neutra, uma vez que, em ratinhos se encontra uma leucina nesta posição. Esta variante foi também encontrada nos doentes estudados no presente trabalho, para os quais se obteve uma frequência alélica de 6,3%.

Ainda em doentes com HPFH, foi descrito numa família da Sardenha a presença de duas mutações no gene *KLF1*, uma *nonsense* (S270X) e uma *missense* (K332Q). Estes autores descreveram, pela primeira vez, a associação de níveis elevados de protoporfirina de zinco com mutações no gene *KLF1*, o que confirma a importância deste gene no controlo da eritropoiese e na biossíntese do heme (Satta *et al.*, 2011).

Recentemente, num estudo em duas crianças da China com anemia refratária Huang *et al.* (2015), determinaram através da técnica de NGS, a presença de uma mutação *frameshift* (c.525_526insCGGCGCC) nas duas crianças e uma mutação *missense* diferente em cada uma delas (p.Ala298Pro e p.Pro338Ser) (Huang *et al.*, 2015).

Num estudo em doentes com SCA foi identificada uma deleção no gene *KLF1* (c.914-4_914-1delCTAG) num doente com HbF elevada de 20,3%. Esta deleção de 4 pb está localizada no início do segundo domínio *zinc finger* no exão 3 e parece alterar o *splicing* (Gallienne, *et al.*, 2012).

No nosso estudo é de realçar que a única criança que não era homozigótica para a mutação que causa a drepanocitose, ou seja, era HbS- β^0 -talassémia, não apresentava variantes para o gene *KLF1*, e era negativa para todos os SNPs estudados, incluindo o *XmnI*. Esta criança tinha um valor de hemoglobina fetal relativamente baixo (5,3%), o que pode estar de acordo com o facto de não ter as variantes estudadas, ou também possivelmente devido a ter uma deleção de 3,7 Kb no gene da α -globina. Satta *et al.* (2011) também associaram níveis baixos de HbF a um indivíduo que apesar de ter variantes no gene *KLF1*, apresentava a deleção no gene da α -globina. Concluímos, pelo tipo e localização das variantes encontradas no gene *KLF1*, que estas poderiam afetar o nível de HbF e, provavelmente, não estariam associadas a qualquer outra patologia.

IV.5.5. Estudo de associação das variantes no gene *KLF1* com o nível de HbF

Para o estudo de associação entre o genótipo relacionado com *KLF1* e o nível de HbF (fenótipo), foram consideradas as 3 mutações *missense* localizadas no exão 2 do referido gene, uma vez que todas pertencem à classe 1, são neutras, ou afetam ligeiramente a expressão do gene. Cada doente foi classificado segundo a presença (+) ou ausência (-) de cada uma dessas variantes em homozigotia ou em heterozigotia. A Tabela IV.11 resume as várias combinações possíveis, o número de doentes que as apresentam, bem como a média e o desvio padrão dos respetivos valores de HbF neste estudo e num estudo de Ngo *et al.* (2013) equivalente.

Tabela IV.11: Mutações *missense* no exão 2 do gene *KLF1* e respetivos valores de HbF

rs112631212 T>G (M39L)		rs2072597A>G (S102P)		rs2072596A>G (F182L)		Nº de doentes	HbF (%)	in Ngo, <i>et al.</i> , 2013	
Homo	Hetero	Homo	Hetero	Homo	Hetero			Nº de doentes	HbF (%)
-	-	+	-	-	-	7	10,1±6,6	2	15,3±5,2
-	-	-	+	-	-	21	10,1±6,6	17	16,3±7,2
-	-	+	-	+	-	2	5,0	0	-
-	+	-	+	-	-	3	10,0±6,2	3	9,3±0,6
-	+	+	-	-	-	3	9,8±6,7	0	-
-	-	-	+	-	+	1	14,0	2	26,1±4,0
-	-	+	-	-	+	2	4,2	1	23
-	-	-	-	-	-	19	11,3±7,3	18	15,6±5,1

+ presença; - ausência

Ngo *et al.* (2013) realizaram um estudo semelhante ao nosso, de associação das referidas mutações no gene *KLF1* e o nível de HbF, em indivíduos com anemia das células falciformes com origem Arábia-Indiana. Através da análise da Tabela IV.11 pode verificar-se que, para as várias combinações possíveis, os indivíduos incluídos no nosso estudo apresentavam, na generalidade, valores inferiores de HbF, o que se pode dever ao facto da nossa população ser de origem Africana. É sabido que os doentes com drepanocitose de origem Arábia-Indiana têm, tendencialmente, níveis de HbF mais elevados que os doentes Africanos (Ngo, *et al.*, 2013).

Em relação ao rs3817621, na região do promotor c.-251C>G, não se encontraram diferenças entre os níveis de HbF (teste Mann-Whitney $p=0,538$), das amostras que apresentavam homo/heterozigotia ($n=23$; HbF: $10,1\pm6,6$) para este SNP e as amostras que não apresentavam este SNP ($n=32$; HbF: $10,9\pm6,9$). Estes resultados estão de acordo com os encontrados por Ngo *et al.* (2013). Concluímos então que são variantes neutras e que, provavelmente, não afetarão a expressão do gene.

IV.5.6. Análise *in silico* através do software PolyPhen-2, das consequências na expressão gênica da nova variante identificada

O PolyPhen-2 (Polymorphism Phenotyping2) é uma ferramenta informática (<http://genetics.bwh.harvard.edu/pph2/>) que permite prever a influência da alteração, ou seja, o seu impacto na patogenicidade da proteína. Assim, foi analisada a alteração identificada que não se encontra descrita, Q342H (rs146658904C>G). Para avaliar o grau de conservação entre espécies desta proteína, as sequências de aminoácidos das proteínas KLF1 de várias espécies foram alinhadas e comparadas, utilizando este programa. Neste alinhamento foi ainda incluída a sequência peptídica humana, tendo-se verificado que o aminoácido nativo (Q342) é altamente conservado entre espécies (Figura IV.21).

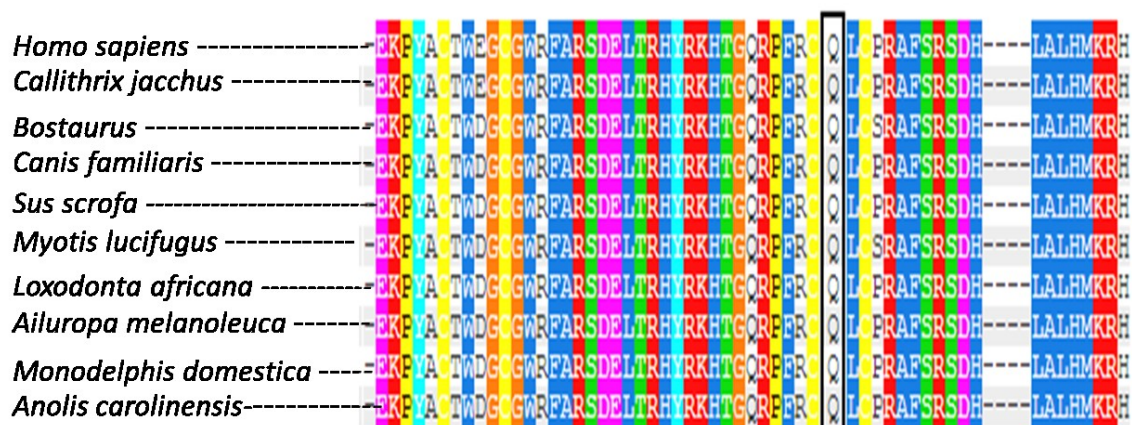


Figura IV.21: Alinhamento da sequência peptídica da proteína KLF1. *Homo sapiens* (Homem); *Callithrix jacchus* (Sagui-de-Tufos-Branços, espécie de macaco); *Bostaurus* (Bovino); *Canis familiaris* (Cão); *Sus scrofa* (Porco); *Myotis lucifugus* (Morcego); *Loxodonta africana* (Elefante); *Ailuropa melanoleuca* (Panda); *Monodelphis domestica* (Gênero de marsupial da família Didelphidae); *Anolis carolinensis* (Camaleão). No retângulo a preto está representado o aminoácido onde se encontrou a alteração no gene *KLF1*.

O programa *PolyPhen-2* foi ainda utilizado para calcular os modelos *HumDiv* e *HumVar* para prever a patogenicidade da proteína. O modelo *HumDiv* é utilizado preferencialmente na avaliação do estudo de alelos raros, mapas de densidade de regiões identificados por estudos de associação de genótipos e na análise da seleção natural. O modelo *HumVar* é utilizado preferencialmente no diagnóstico de doenças Mendelianas. Neste caso pretendeu-se averiguar em que medida a alteração Q342H poderá interferir na estrutura final de KLF1. No modelo *HumDiv* esta alteração apresentou um *score* 0,013 e no modelo *HumVar* de 0,063, indicando, ambos, uma previsão de impacto benigno na alteração da proteína (Figura IV.22).

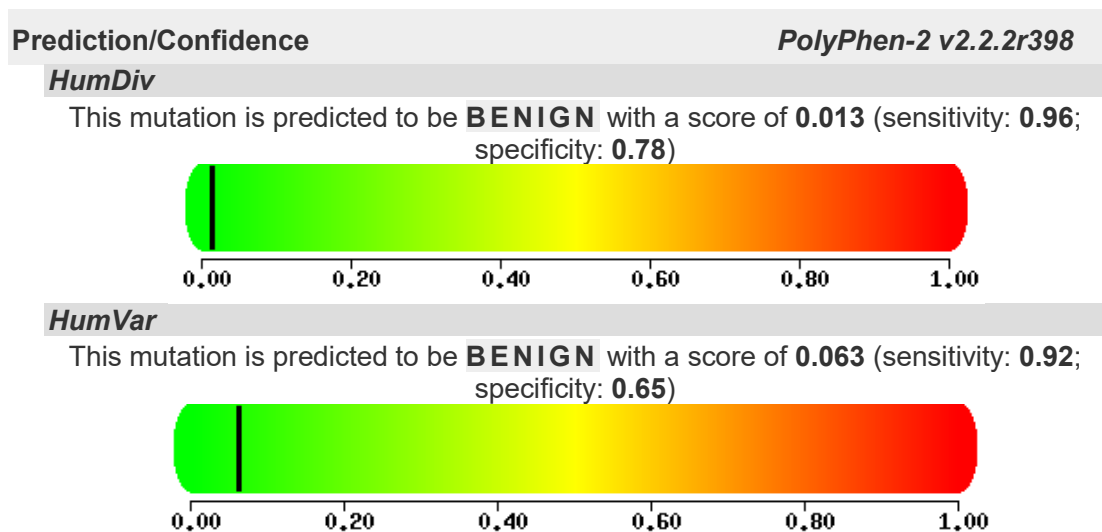


Figura IV.22: Resultado da análise do modelo *HumDiv* e *HumVar* através do programa *PolyPhen-2*. A linha a preto representa o *score* da variante identificada e como se encontra na região verde significa que a variante Q342H não é patogénica.

IV.6. As diferenças nas frequências das variantes no gene *KLF1* entre populações e patologias

O gene *KLF1* foi inicialmente descrito como um ativador da expressão do gene da β -globina. Contudo, o *KLF1* regula diretamente a expressão de vários genes de forma a coordenar a produção de uma molécula de hemoglobina funcional, regulando a produção não só de β -globina, como também de α -globina (Tallack & Perkins, 2010). As mutações no gene *KLF1* foram, pela primeira vez, descritas em 2008 por Singleton *et al.*, as quais foram associadas ao fenótipo In(Lu) (Singleton, Burton, Green, Brady & Anstee, 2008). Os antígenos do grupo de sangue Lutheran são glicoproteínas localizadas na membrana

celular do eritrócito codificadas pelo gene *BCAM*; quando estas não estão presentes, ou são fracamente expressas, dão origem ao fenótipo Lu(a-b-) (Borg, Patrinos, Felice & Philipsen, 2011). As alterações genéticas que podem dar origem a este fenótipo são de três tipos: a) mutações recessivas no gene *BCAM*; b) mutações ligadas ao cromossoma X; c) mutações no inibidor dominante de Lutheran ou In(Lu) (Waye & Eng, 2015). No entanto, este gene tem sido também associado a aumento de HbF e HbA₂, HPFH e anemia deseritropoiética congénita (Borg, *et al.*, 2011).

Gallienne *et al.* (2012), encontraram mutações no gene *KLF1* em 11 dos 131 (8,4%), dos doentes propostos ao laboratório para estudo de hemoglobinopatias, que apresentavam valores elevados de HbF e com diferentes origens étnicas (África, Índia e Sudoeste Asiático) (Gallienne *et al.*, 2012). Todas as mutações encontradas foram em heterozigotia, o que mostra que *in vivo* uma alteração em apenas um alelo é o suficiente para aumentar os níveis de HbF. Em contrapartida, não encontraram mutações em *KLF1* na população controlo com HbF normal. Estes resultados indicam que mutações no *KLF1* podem ter uma contribuição importante na variação de HbF em regiões onde a malária é ou foi prevalente e onde as hemoglobinopatias são comuns (Gallienne, *et al.*, 2012). Em contraste, no nosso estudo de pesquisa de variantes no gene *KLF1*, em doentes africanos com anemia das células falciformes, um número bastante elevado de doentes, 48 em 58 (82,8%) apresentava, pelo menos, uma das variantes encontradas, enquanto 10 em 58 (17,2%) não apresentava nenhuma variante. Esta grande diferença, entre os dois estudos, na proporção de doentes com variantes em *KLF1* (8,4 vs 82,8%) poderá ser devida a: 1) diferenças étnicas entre as populações estudadas (Africanos vs Indianos e do Sudoeste Asiático); 2) diferenças entre as patologias envolvidas (SCA vs outras hemoglobinopatias), ou 3) falha na deteção por erros metodológicos (falsos negativos) no caso do estudo de Gallienne, *et al.*, 2012.

IV.7. A ação de efeito aditivo das variantes sobre o nível de HbF

Atendendo a que, sempre que possível, o indivíduo deverá ser avaliado como um todo, analisámos o efeito conjunto das variantes estudadas nos vários genes sobre o nível de HbF. Assim, na Tabela IV.12 resume-se a combinação das variantes encontradas no gene *KLF1* com os vários SNPs estudados neste trabalho, em *HBG2*, *BCL11A*, *HBS1L-MYB* e *MYB*, bem como a média e o desvio padrão dos respetivos valores de HbF.

Tabela IV.12: Níveis de HbF dos indivíduos com anemia das células falciformes, de acordo com a presença de variantes em *KLF1* e outros polimorfismos genéticos nos genes *HBG2*, *BCL11A*, *HBS1L-MYB* e *MYB*

<i>KLF1</i>	<i>HBG2</i> <i>XmnI</i> (rs7482144)	<i>BCL11A</i> (rs4671393)	n	HbF (%)
		<i>HBS1L-MYB</i> (rs4895441)		
		<i>MYB</i> (rs9399137)		
-	-/-	-	3	9,0±3,3
		+	4	14,1±9,7
		+	1	9,5
	+/-	-	1	9,2
		+	1	16,0
+	-/-	-	17	7,6±6,2
		+	16	9,0±6,3
		-	2	14,0
	+/-	+	4	15,9±7,0
		-	1	14,6
	+/+	+	8	14,9±5,2

+ presença; - ausência

Através dos dados obtidos traçou-se o gráfico representado na Figura IV.23.

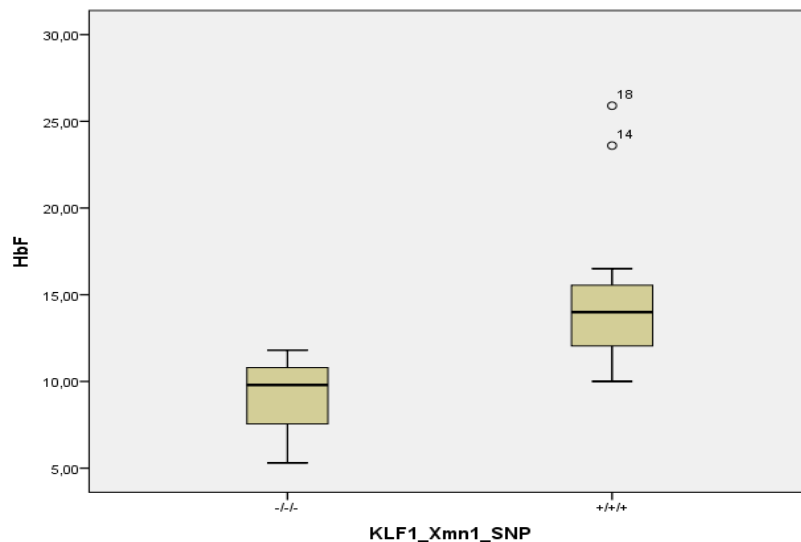


Figura IV.23: Box-Plot das amostras para a combinação das variantes para o gene *KLF1*, *HBG2* (*XmnI*; rs7482144), *BCL11A* (rs4671393), *HBS1L-MYB* (rs4895441) e *MYB* (rs9399137). O sinal – significa que não tem a variante, e o sinal + significa que apresenta a variante. Assim -/-/- significa que não tem qualquer destas variantes, e +/+/+ significa que tem as diferentes variantes aqui representadas.

Através da análise estatística, entre o grupo de doentes que não apresenta as variantes acima definidas e o grupo que apresenta as respetivas variantes, pode aplicar-se o teste de Kruskal-Wallis e a um nível de confiança de 95% obtém-se um $p=0,021$, pelo que concluímos que existem diferenças estatisticamente significativas. A combinação das variantes do *KLF1* com o *XmnI* (*HBG2*; rs7482144) e com os SNPs do *BCL11A* e *HBSIL-MYB* está associada a níveis mais elevados de HbF (Figura IV.23). Estes resultados estão de acordo com os encontrados, muito recentemente, por Tepakhan *et al.* (2016), num estudo semelhante em que se estudou as mesmas variantes, mas em doentes com hemoglobina E (Tepakhan *et al.*, 2016). Podemos então concluir que *KLF1* é um importante fator genético associado ao aumento dos níveis de HbF, especialmente quando em combinação com outros modificadores genéticos. A ação conjunta das variantes parece ter um efeito aditivo ou sinérgico sobre o nível de HbF, e pode explicar a variação fenotípica dos níveis de HbF nestes doentes com anemia das células falciformes.

IV.8. Estudos de associação entre os genótipos das variantes do *KLF1* e o risco de AVC

Com o objetivo de avaliar a possível associação de alguma das variantes identificadas ao risco de AVC, utilizando NGS do gene *KLF1*, foram realizadas várias análises e tratamento estatístico, nas quais se analisou os genótipos de cada variante, para cada um dos três grupos em estudo (Controlo, Risco e AVC). Em seguida, analisou-se a variação dos genótipos do grupo controlo, em função do grupo de risco associado ao grupo AVC. Esta análise foi efetuada apenas para as variantes identificadas, que apresentaram uma frequência alélica superior a 4%. Os resultados encontram-se resumidos na Tabela IV.13.

Tabela IV.13: Estudo estatístico para as diferentes variantes genéticas identificadas no gene *KLF1*, mostrando as várias combinações dos genótipos com os vários grupos estudados

Gene	Variante	Genótipo (valor p)	
		Controlo vs Risco vs AVC	Controlo vs AVC+Risco
<i>KLF1</i>	rs3817621	0,245	0,285
	rs112631212	0,168	0,472
	rs2072597	0,310	0,344
	rs2072596	0,488	0,352
	rs16978757	0,170	0,116
	rs16978754	0,170	0,116

Através da análise da Tabela IV.13 verifica-se que não se encontrou associação entre as diferentes variantes do gene *KLF1* e o risco de AVC.

Como o AVC é uma doença multifatorial, não é possível associar a uma só variante genética, ou seja, terá de se associar a vários genes, diferentes variantes e diversos fatores ambientais, que vão contribuir para o desenvolvimento da doença.

CAPITULO V
CONCLUSÕES E PERSPETIVAS FUTURAS

V – CONCLUSÕES E PERSPETIVAS FUTURAS

Este estudo permitiu avaliar o efeito do nível de HbF no desenvolvimento de AVC num grupo de crianças africanas residentes em Portugal, com anemia das células falciformes. Observou-se que um nível baixo de HbF é um fator de risco para o desenvolvimento deste sub-fenótipo da doença, uma vez que se encontraram diferenças estatisticamente significativas ($p=0,005$) nos níveis deste parâmetro entre os diferentes grupos de doentes (Controlo / Risco / AVC). Este conceito de que a baixa HbF é um fator de risco para AVC na SCA já tinha sido proposto por alguns autores (Platt, *et al.*, 1994; Alsultan, *et al.*, 2012) e é aqui confirmado. Reforçamos, assim, o conceito de que o nível de HbF é um fator muito importante na modulação das manifestações clínicas desta patologia.

Como é sabido que a HbF, é ela própria, modulada por fatores genéticos, avaliou-se o efeito de diferentes variantes genéticas, tanto globínicas (no gene *HBG2*), como não globínicas (nos genes *BCL11A*, *HBSIL-MYB*, *MYB* e *KLF1*) sobre o nível de HbF e também sobre o desenvolvimento de AVC. Concluiu-se que as variantes dos genes *HBG2* e *BCL11A* parecem exercer uma ação mais forte sobre a regulação do nível de HbF, do que variantes na região intergénica *HBSIL-MYB* e no gene *MYB*. Assim, observámos uma tendência do genótipo TT do rs7482144 do gene do *HBG2* para influenciar o aumento dos níveis de HbF ($p=0,096$), a qual se verificou ser estatisticamente significativa quando se agrupam os doentes com os genótipos TT+CT ($p=0,031$). Ou seja, é suficiente a presença de um alelo variante T para fazer subir o nível de HbF. Estes resultados estão de acordo com os descritos na literatura (Gilman & Huisman, 1984; Cardoso, *et al.*, 2014). Quanto ao SNP rs11886868 do gene *BCL11A*, verifica-se uma tendência para valores mais elevados de HbF no genótipo CC, embora não estatisticamente significativa ($p=0,136$). Do mesmo modo, a variante do SNP rs4671393 no mesmo gene, ou seja, o alelo A, está tendencialmente relacionado com o aumento de HbF, embora não estatisticamente significativo ($p=0,057$). Estes resultados estão de acordo com os reportados por Wonkam *et al.* (2014), numa população com SCA nos Camarões, embora neste caso, tenham sido observadas diferenças estatisticamente significativas (Wonkam *et al.*, 2014).

Concluiu-se também, que a metodologia utilizada para o estudo do gene *KLF1* e regiões regulatórias proximais, por PCR-Longo (num total de 3,2 Kb) seguido de NGS, se mostrou bastante adequada. Verificou-se que 82,8% dos nossos doentes com SCA (48 em 58) apresentavam, pelo menos, uma variante neste gene. Foram identificadas 11 variantes

diferentes localizadas ao longo do gene, sendo 4 *missenses*, 2 intrónicas, e 5 localizam-se em regiões regulatórias 5' - ou 3'-UTR.

As frequências alélicas encontradas para estas variantes são semelhantes às descritas para a população africana em geral, em base de dados públicas, exceto para duas das variantes: rs3817621G>C e rs2072597A>G. Nestas, a frequência alélica encontrada foi superior à descrita, respetivamente, 29,2% *versus* 19% e 55,2% *versus* 44%. Estes valores, a serem confirmados, podem sugerir que estas duas variantes sejam benéficas para esta patologia e, portanto, ao longo dos anos tenham sofrido uma seleção positiva. Quanto à variante nova no gene *KLF1*, encontrada em heterozigotia num dos doentes estudados, (Q342H; rs146658904C>G), os estudos realizados *in silico* indicaram uma previsão de impacto benigno na alteração da proteína.

Observou-se que o efeito de modulação sobre o nível de HbF é muito mais notório considerando o conjunto das variantes, e não apenas cada uma isoladamente (efeito aditivo ou sinérgico). Assim, observou-se que a combinação das variantes do gene *KLF1* com o SNP em *HBG2* e com os SNPs do *BCL11A* e *HBSIL-MYB* está estatisticamente associada a níveis mais elevados de HbF (**p=0,021**). Estes resultados estão de acordo com os encontrados por Tepakhan *et al.* (2016) num estudo semelhante, em que se estudou as mesmas variantes, mas em doentes com outra hemoglobinopatia (HbE). O *KLF1* é um importante fator genético associado ao aumento dos níveis de HbF e, quando em combinação com outros modificadores genéticos, pode explicar a variação fenotípica dos níveis de HbF nestes doentes com anemia das células falciformes.

Embora se tenha observado que a associação das diferentes variantes genéticas no seu conjunto são nitidamente fatores modificadores do nível de HbF (genótipo → fenótipo hematológico/bioquímico intermédio) nos doentes com SCA e que, por outro lado, os níveis baixos de HbF estão correlacionados com o desenvolvimento de AVC (fenótipo hematológico/bioquímico intermédio → fenótipo clínico final), não foi possível correlacionar diretamente a ausência das variantes genéticas com o AVC (genótipo_ fenótipo clínico final). Isto deveu-se, provavelmente, ao número baixo da amostra quando subdividida, atendendo aos vários parâmetros analisados e também, provavelmente, ao facto de que o AVC, para além de ser modulado pelo nível de HbF, também o ser por vários outros fatores.

É também de concluir que nos foi possível estabelecer um perfil genético de risco relacionado com níveis baixos de HbF na SCA: ausência de variantes no gene *KLF1*; *HBG2*_rs7482144_CC; *BCL11A*_rs11886868_TT; *BCL11A*_rs4671393_GG; *HBSIL-MYB*_rs4895441_AA; *MYB*_rs9399137_TT; *MYB*_rs9389268_AA.

Podemos ainda concluir que a identificação e caracterização dos modificadores genéticos estudados é muito importante nestes doentes pois, como estão envolvidos em mecanismos fisiopatológicos subjacentes às vasculopatias cerebrais, poderão vir a constituir alvos de novas abordagens terapêuticas, incluindo a terapia génica. Esta é uma área de grande investigação a nível mundial.

Ainda como perspetiva futura, para continuação deste trabalho, está em projeto a confirmação destes resultados numa população com SCA equivalente à nossa, mas residente no estado de São Paulo, Brasil. Está também em aberto o estudo de outros genes modificadores de HbF e de vasculopatia cerebral. Por exemplo, Sheehan *et al.* (2014), associaram uma variante codificante no fator de transcrição *Spalt-like* (*SALL2 – Spalt-like transcription factor*) a valores elevados de HbF (Sheehan, *et al.*, 2014). Este pode ser um gene candidato a futuras investigações na população com SCA. Outros fatores de transcrição envolvidos no silenciamento dos genes das globinas nas células eritróides adultas, tais como, *SOX6*, *GATA1* e *LSU1*, deverão também vir a ser considerados.

Em termos metodológicos propõe-se alargar a aplicação da metodologia de NGS aos estudos das diferentes variantes dos diversos genes através da construção de um painel de genes, o que tornará o processo de caracterização das variantes muito mais rápido, requererá menor quantidade de DNA e será possível analisar um elevado número de amostras simultaneamente.

CAPITULO VI
BIBLIOGRAFIA

VI – BIBLIOGRAFIA

Adams, R. J. (2007). Big Strokes in Small Persons. *Neurological Review*, 64 (11), 1567-1574.

Adewoyin, A. S. (2015). Management of sickle cell disease: a review for physician education in Nigeria (Sub-Saharan Africa). *Hindawi Publishing Corporation Anemia*, 2015, ID 791498, 21 páginas. doi: 10.1155/2015/791498

Akinsheye, I., Solovieff, N., Ngo, D., Malek, A., Sebastiani, P., Steinberg, M. H. & Chui, D. H. K. (2011). Fetal hemoglobin in sickle cell anemia: molecular characterization of the unusually high fetal hemoglobin phenotype in African Americans. *American Journal of Hematology*, 87 (2), 217-219.

Alsultan, A., Aleem, A., Ghabbour, H., AlGahtani, F. H., Al-Shehri, A., Osman, M. E., ..., Al-Momen, A. M. (2012). Sickle cell disease subphenotypes in patients from Southwestern Province of Saudi Arabia. *J Pediatr Hematol Oncol.*, 34(2), 79-84. doi: 10.1097/MPH.0b013e3182422844.

Arnaud, L., Saison, C., Helias, V., Lucien, N., Steschenko, D., Giarratana, M-C., ..., Cartron, J-P. (2010). A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *The American Journal of Human Genetics*, 87, 721-727.

Bauer, D. E., Kamran, S. C. & Orkin, S. H. (2012). Reawakening fetal hemoglobin: prospects for new therapies for the β -globin disorders. *Blood*, 120 (15), 2945-2953. doi: 10.1182/blood-2012-06-292078

Bordin, S., Martins, J. T., Gonçalves, M. S., Melo, M. B., Saad, S. T. O. & Costa, F. F. (1998). Haplotype analysis and $A\gamma$ gene polymorphism associated with the Brazilian type of hereditary persistence of fetal hemoglobin. *American Journal of Hematology*, 58, 49-54.

Borg, J., Papadopoulos, P., Georgitsi, M., Gutiérrez, L., Grech, G., Fanis, P., ..., Philipsen, S. (2010). Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nature Genetics*, 42 (9), 801-805. doi: 10.1038/ng.630.

Borg, J., Patrinos, G. P., Felice, A. E., & Philipsen, S. (2011). Erythroid phenotypes associated with *KLF1* mutations. *Haematologica (Editorials & Perspectives)*, 96 (5), 635-638.

Cao, A., Moi, P. & Galanello, R. (2011). Recent advances in β .thalassemias. *Pediatric Reports*, 3 (17), 65-78.

Cardoso, G. L., Diniz, I. G., da Silva, A. N. L. M., Cunha, D. A., Junior, J. S. S., Uchôa, C. T. C., ..., Guerreiro, J. F. (2014). DNA polymorphisms at *BCL11A*, *HBS1L-MYB* and *Xmn1-HBG2* site loci associated with fetal hemoglobin levels in sickle cell anemia patients from Northern Brazil. *Blood Cells, Molecules and Diseases*, 53, 176-179.

Coelho, A., Dias, A., Moraes, A., Nunes, B., Ferreira, E., Picanço, I., Faustino, P. e Lavinha, J. (2013). Genetic variation in *CD36*, *HBA*, *NOS3* and *VCAM1* is associated with chronic haemolysis level in sickle cell anaemia: a longitudinal study. *European Journal of Haematology*, 92, 237-243.

Costa, D., Capuano, M., Sommese, L. & Napoli, C. (2015). Impact of epigenetic mechanisms on therapeutic approaches of hemoglobinopathies. *Blood Cells, Molecules and Diseases*, 55, 95-100.

Danjou, F., Anni, F. & Galanello, R. (2011). Beta-thalassemia: from genotype to phenotype. *Haematologica*, 96 (11), 1573- 1575. doi: 10.3324/haematol.2011.055962

Danjou, F., Anni, F., Perseu, L., Satta, S., Dessi, C., Lai, M. E., ..., Galanello, R. (2012). Genetic modifiers of β -thalassemia and critical severity as assessed by age at first transfusion. *Haematologica*, 97 (7), 989-993.

Domingos, I. F., Falcão, D. A., Hatzlhofer, B. L., Cunha, A. F., Santos, M. N., Albuquerque, D. M., ..., Bezerra, M. A. (2014). Influence of the β s haplotype and α -thalassemia on stroke development in a Brazilian population with sickle cell anaemia. *Ann Hematol*, 93 (7), 1123-1129. doi: 10.1007/s00277-014-2016-1.

Driss, A., Asare, K. O., Hibbert, J. M., Gee, B. E., Adamkiewicz, T. V. & Stiles, J. K. (2009). Sickle cell disease in the post genomic era: a monogenic disease with a polygenic phenotype. *Genomics Insights*, 2, 23-48.

Elmariah, H., Garrett, M. E., de Castro, L., Jonassaint, J., Ataga, K. I., Eckman, J., ..., Telen, M. J. (2014). Factors Associated with Survival in a Contemporary Adult Sickle Cell Disease Cohort. *Am J Hematol*, 89 (5), 530–535. doi:10.1002/ajh.23683.

Finotti, A., Breda, L., Lederer, C. W., Bianchi, N., Zuccato, C., Kleanthous, M., ..., Gambari, R. (2015). Recent trends in the gene therapy of β -thalassemia. *Journal of Blood Medicine*, 6, 69-85.

Frenette, P. S. & Atweh, G. F. (2007). Sickle cell disease: old discoveries, new concepts, and future promise. *The Journal of Clinical Investigation*, 117 (4), 850-858.

Friedrich, J. R., Sheehan, V., Flanagan, J. M., Baldan, A., Summavrell, C. C. G., Bittar, C. M., ..., Silla, L. M. R. (2016). The role of *BCL11A* and *HMIP-2* polymorphisms on endogenous and hydroxyurea induced levels of fetal hemoglobin in sickle cell anemia patients from southern Brazil. *Blood Cells, Molecules and Diseases*, 62, 32-37.

Galanello, R., Sanna, S., Perseu, L., Sollaino, M. C., Satta, S., Lai, M. E., ..., Cao, A. (2009). Amelioration of Sardinian β^0 thalassemia by genetic modifiers. *Blood*, 114 (18), 3935-3937. doi: 10.1182/blood-2009-04-217901

Gallienne, A. E., Dréau, H. M. P., Schuh, A., Old, J. M. & Henderson, S. (2012). Ten novel mutations in the erythroid transcription factor *KLF1* gene associated with increased fetal hemoglobin levels in adults. *Haematologica*, 97 (2), 340-343.

Gilman, J. G. & Huisman, T. H. (1984). Two independent genetic factors in the β -globin gene cluster are associated with high $G\gamma$ -levels in the HbF of SS patients. *Blood*, 64 (2), 452-457.

Graffelman, J. & Weir, B.S. (2016). Testing for Hardy–Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity* advance online publication, 20, 1-11. doi:10.1038/hdy20

Huang, J., Zhang, X., Liu, D., Wei, X., Shang, X., Xiong, F., ..., Xu, X. (2015). Compound heterozygosity for *KLF1* mutations is associated with microcytic hypochromic anemia and increased fetal hemoglobin. *European Journal of Human Genetics*, 23, 1341-1348.

Kato, G. J., Gladwin, M. T. & Steinberg, M. H. (2007). Deconstructing sickle cell disease: reappraisal of the role of hemolysis in the development of clinical subphenotypes. *Blood*, 21 (1), 37-47.

Kiefer, C. M., Hou, C., Little, J. & Dean, A. (2008). Epigenetics of β -globin gene regulation. *Mutat Res*, 647(1-2): 68–76. doi:10.1016/j.mrfmmm.2008.07.014.

Lettre, G. & Bauer, D. E. (2016). Fetal haemoglobin in sickle-cell disease: from genetic epidemiology to new therapeutic strategies. *Lancet*, 387, 2554-2564.

Lettre, G., Sankaran, V. G., Bezerra, M. A. C., Araújo, A. S., Uda, M., Sanna, S., ..., Orkin, S. H. (2008). DNA polymorphisms at the *BCL11A*, *HBSIL-MYB*, and β -globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *PNAS*, 105 (33), 11869-11874.

Lowrey, C. H. (2016). Down the repressors! Up the fetal hemoglobin!. *Blood*, 127 (11), 1384-1385. doi: 10.1182/blood-2016-01-689018

Manwani, D. & Bieker, J. J. (2014). KLF1: when less is more. *Blood*, 124 (5), 672-673. doi: 10.1182/blood-2014-05-576967

Miller, B. A., Salameh, M., Ahmed, M., Wainscoat, J., Antognetti, G., Orkin, S., ..., Nathan, D. G. (1986). High fetal hemoglobin production in sickle cell anemia in the eastern province of Saudi Arabia is genetically determined. *Blood*, 67 (5), 1404-1410.

Ngo, D. A. & Steinberg, M. H. (2015). Genomic approaches to identifying targets for treating β hemoglobinopathies. *BMC Medical Genomics*, 8:44. doi: 10.1186/s12920-015-0120-2

Ngo, D., Bae, H., Steinberg, M.H., Sebastiani, P., Solovieff, N., Baldwin, C.T., ... Alsultan, A. (2013). Fetal Hemoglobin in Sickle Cell Anemia: Genetic Studies of the Arab-Indian Haplotype. *Blood Cells Mol Dis.*, 51(1), 22–26, doi:10.1016/j.bcmd.2012.12.005.

Paciaroni, K., Lucarelli, G., Martelli, F., Migiliaccio, A. R., von Lindern, M., Borg, J., ..., Phillipsen, S. (2014). Transfusion-independent β^0 -thalassemia after bone marrow transplantation failure: proposed involvement of high parental HbF and an epigenetic mechanism. *Am J Blood Res*, 4 (1), 27-32.

Pakbaz, Z. & Wun, T. (2014). Role of the hemostatic system on SCA pathophysiology and potential therapeutics. *Hematol Oncol Clin North Am.*, 28 (2), 355-374. doi:10.1016/j.hoc.2013.11.011.

Pereira, C., Relvas, L., Bento, C., Abade, A., Ribeiro, M. L. & Manco, L. (2015). Polymorphic variations influencing fetal hemoglobin levels: association study in beta-thalassemia carriers and in normal individuals of Portuguese origin. *Blood Cells, Molecules and Diseases*, 54, 315-320.

Perkins, A., Xu, X., Higgs, D. R., Patrinos, G. P., Arnaud, L., Bieker, J. J. & Phillipsen, S. (2016). Krüppeling erythropoiesis: an unexpected broad spectrum of human

red blood cell disorders due to *KLF1* variants. *Blood*, 127 (15), 1856-1862. doi: 10.1182/blood-2016-01-694331

Platt, O. S., Brambilla, D. J., Rosse, W. F., Milner, P. F., Castro, O., Steinberg, M. H. & Klug, P. P. (1994). Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N Engl J Med.*, 330(23), 1639-1644.

Pule, G. D., Bitoungui, V. J. N., Chemegni, B. C., Kengne, A. P., Antonarakis, S. & Wonkam A. (2015). Association between variants at *BCL11A* erythroid-specific enhancer and fetal hemoglobin levels among sickle cell disease patients in Cameroon: implications for future therapeutic interventions. *OMICS: A Journal of Integrative Biology*, 19 (10), 627-631. doi: 10.1089/omi.2015.0124

Razin, S. V., Ulianov, S. V., Ioudinkova, E. S., Gushchanskaya E. S., Gavrilov A. A. & Iarovaia, O. V. (2012). Domains of α - and β -globin genes in the context of the structural-functional organization of the eukaryotic genome. *Biochemistry (Mosc)*, 77(13):1409-1423. doi: 10.1134/S0006297912130019

Saiki, R. K., Chang, C.A., Levenson, C. H., Warren, T. C., Boehm, C. D., Kazazian, H. H. & Erlich, H. A. (1988). Diagnosis of sickle cell anemia and beta-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. *N Engl J Med.*, 319 (9), 537-541. doi: 10.1056/NEJM198809013190903

Saki, N., Abroun, S., Soleimani, M., Kavianpour, M., Shahjahani, M., Mohammadi-Asl, J. & Hajizamani, S. (2016). MicroRNA expression in β -thalassemia and sickle cell disease: a role in the induction of fetal hemoglobin. *Cell Journal*, 17 (4), 583-592.

Sankaran, V. G. & Orkin, S. H. (2013). The Switch from Fetal to Adult Hemoglobin. *Cold Spring Harbor Perspectives in Medicine*, 3: a011643. doi: 10.1101/cshperspect.a011643

Sankaran, V. G. (2011). Targeted therapeutic strategies for fetal hemoglobin induction. *American Society of Hematology*, 10 (1), 459-465. doi:10.1182/asheducation-2011.1.459

Satta, S., Perseu, L., Moi, P., Asunis, I., Cabriolu, A., Maccioni, L., ..., Galanello, R. (2011). Compound heterozygosity for *KLF1* mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin. *Haematologica*, 96 (5), 767-770.

Schechter, A. N., (2008). Hemoglobin research and the origins of molecular medicine. *Blood*, 112 (10), 3927-3938. doi: 10.1182/blood-2008-04-078188

Schnog, J. B., Duits, A. J., Luskiet, F. A. J., ten Cate, H. & Rojer, R. A. (2004). Sickle cell disease; a general overview. *The Netherlands Journal of Medicine*, 62 (10), 364-374.

Scheps, K. G., Francipane, L., Nash, A., Cerrone, G. E., Copelli, S. B. & Varela, V. (2015). Bases moleculares de alfa-talasemia en la argentina. *Medicina*, 75, 81-86.

Sebastiani, P., Wang, L., Nolan, V. G., Melista, E., Ma, Q., Baldwin, C. T. & Steinberg, M. H. (2008). Fetal hemoglobin in sickle cell anemia: Bayesian modeling of genetic associations. *American Journal of Hematology*, 83, 189-195. doi: 10.1002/ajh.21048

Sheehan, V. A., Crosby, J. R., Sabo, A., Mortier, N. A., Howard, T. A., Muzny, D. M., ..., Flanagan, J. M. (2014). Whole exome sequencing identifies novel genes for fetal hemoglobin response to hydroxyurea in children with sickle cell anemia. *PLoS One*, 9 (10), e110740. doi: 10.1371/journal.pone.0110740

Siatecka, M. & Bieker, J. J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood*, 118 (8), 2044-2054. doi: 10.1182/blood-2011-03-331371

Singleton, B. K., Burton, N. M., Green, C. Brady, R. L. & Anstee, D. J. (2008). Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood*, 112 (5), 2081-2088.

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Experimental Hematology*, 33 (3), 259 – 271.

Steinberg, M. H., Barton, F., Castro, O., Pegelow, C. H., Ballas, S. K., Kutlar, A., ..., Terrin, M. (2003). Effect of hydroxyurea on mortality and morbidity in adult sickle cell anemia: risks and benefits up to 9 years of treatment. *JAMA*, 289(13):1645-1651.

Steinberg, M.H. (2009). Genetic etiologies for phenotypic diversity in sickle cell anemia. *The Scientific World Journal*, 9, 46– 67. doi: 10.1100/tsw.2009.10

Switzer, J. A., Hess, D. C., Nichols, F. T. & Adams, R. J. (2006). Pathophysiology and treatment of stroke in sickle-cell disease: present and future. *Lancet Neurol.*, 5, 501-512.

Tallack, M. R. & Perkins, A. C. (2010). KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. *Life*, 62 (12), 886-890.

Tepakhan, W., Yamsri, S., Sanchaisuriya, K., Fucharoen, G., Xu, X. & Fucharoen, S. (2016). Nine known and five novel mutations in the erythroid transcription factor *KLF1* gene and phenotypic expression of fetal hemoglobin in hemoglobin E disorder. *Blood Cells, Molecules and Diseases*, 59, 85-91.

Thein, S. L., Menzel, S., Lathrop, M. & Garner, C. (2009). Control of fetal hemoglobin: new insights emerging from genomics and clinical implications. *Human Molecular Genetics*, 18 (2), R216-R223. doi: 10.1093/hmg/ddp401

Theml, H., Diem, H. & Haferlach, T. (2004). *Color atlas of hematology – practical microscopic and clinical diagnosis*. 2ª Edição. New York, EUA: Thieme Flexibook.

Vinjamur, D. S., Alhashem, Y. N., Mohamad, S. F., Amin, P., Williams, D. C. & Lloyd, J. A. (2016). Krüppel-like transcription factor *KLF1* is required for optimal γ - and β -globin expression in human fetal erythroblasts. *PLOS ONE*, 11 (2), e0146802. doi: 10.1371/journal.pone.0146802

Waye, J. S. & Eng, B. (2015). Krüppel-like factor 1: hematologic phenotypes associated with *KLF1* gene mutations. (2015). *International Journal of Laboratory Hematology*, 37 (Suppl. 1), 78-84.

Weatherall, D. J. (2001). Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. *Nat Rev Genet.*, 2(4), 245-55.

Wonkam, A., Bitoungui, V. J. N., Vorster, A. A., Ramesar, R., Cooper, R. S., Tayo, B., Lettre, G. & Ngogang, J. (2014). Association of variants at *BCL11A* and *HBS1L-MYB* with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. *PLos One*, 9 (3), e92506. doi: 10.1371/journal.pone.0092506

Ferramentas bioinformáticas:

<http://genetics.bwh.harvard.edu/pph2/>

<http://www.ensembl.org/index.html>

<https://genome.ucsc.edu/>

<https://www.ncbi.nlm.nih.gov/pubmed>

Tabela A.1: Composição do tampão B utilizado nos PCRs e do tampão de aplicação de azul de bromofenol utilizado na aplicação das amostras nos géis de agarose

Tampão B		Tampão de aplicação nos géis de agarose
Solução Stock (concentração final)	Solução de Trabalho	
KCl₂ 2M - 50 mM Tris-HCl 1,5 M (pH = 8,8) - 100 mM MgCl₂ 1 M - 15 mM Gelatina – 0,1% H₂O – 26,73 mL	Solução Stock – 0,5 mL H₂O - 4,5 mL dNTPs 100mM - 40 µL de cada	Azul de Bromofenol – 0,22 g H₂O - 8,75µL Glicerol - 18,75µL EDTA - 75µL NaOH - até ficar azul

Tabela A.2: Condições de PCR para amplificação do fragmento do gene *HBB* para pesquisa molecular de mutação drepanocítica

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	Tm (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>HBB</i> (cr. 11)	GAG->GTG Mutação no codão 6 (HBB: c.20A>T)	1 (Forward)	5' – ACCTCACCTGTGGAGCCAC- 3'	20	66	Tampão B	22,9	94	5'	390
						Primer 1	0,5	94	45''	
						Primer 67	0,5	65	45'' ×29	
		67 (Reverse)	5'- ACCAGCAGCCCTAAGGGTGGGAAAATACACC – 3'	31	81	Taq	0,1	72	45''	
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.3: Condições da digestão enzimática para pesquisa molecular de mutação drepanocítica

		Mistura da Reação (Por Amostra)		Condições da Reação		Produto de restrição	
Gene	Variante Genética	Reagente	Volume (µL)	T (°C)	Δt	Alelo T (mutação)	Alelo A (normal)
<i>HBB</i> (cr. 11)	G <u>A</u> G->G <u>T</u> G Mutação no codão 6 (HBB: c.20A>T)	<i>Bsu36I</i> (Biolabs)	0,5 (5U)	37	3 horas ou <i>Overnight</i>	(-)	(+)
		H ₂ O	4,5			390 pb	177 pb
		Sequência de reconhecimento			213 pb		
		Produto PCR	10	5' ... C C ▼ T N A G G ... 3' 3' ... G G A N T A C C ... 5'			

Tabela A.4: Condições de PCR para amplificação de fragmento do *cluster* da β-globina para análise do rs7482144 do gene *HBG2*

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	T _m (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>HBG2</i> (cr. 11)	rs7482144	R 160 S (<i>Forward</i>)	5' – GCACTGAAACTGTTGCTTTATAGGAT - 3'	26	54,4	Tampão B	22,9	94	5'	677
						Primer R 160 S	0,5	94	45''	
						Primer R 161 AS	0,5	55	45'' ×30	
		R 161 AS (<i>Reverse</i>)	5' - TGGCGTCTGGACTAGGAGCTTATT - 3'	24	57,4	Taq	0,1	72	45''	
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.5: Condições da digestão enzimática para análise do rs7482144 do gene *HBG2* com *XmnI*

		Mistura da Reação (por amostra)		Condições da Reação		Produto de restrição	
Gene	Variante Genética	Reagente	Volume (µL)	T (°C)	Δt	Alelo T (Variante)	Alelo C (normal)
HBG2 (cr. 11)	rs7482144	XmnI (Biolabs)	0,25 (5U)	37	3 horas ou Overnight	(+)	(-)
		Tampão	1,5			462 pb	677 pb
		H2O	3,25	Sequência de reconhecimento			
		Produto PCR	10	5' ... GAANN▼NN■TC ... 3' 3' ... CTTNN▲NNAAG ... 5'		215 pb	

Tabela A.6: Condições de PCR para amplificação de fragmento do gene *BCL11A* para análise do rs11886868

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	T _m (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>BCL11A</i> (cr. 2)	rs11886868	BCL11A-R (<i>Forward</i>)	5'- GGGCTCCCTGGACTCAAAC TC - 3'	21	59,8	Tampão B	22,9	94	5'	659
						Primer BCL11A-R	0,5	94	45''	
						Primer BCL11A-F	0,5	62	45'' ×30	
		BCL11A-F (<i>Reverse</i>)	5' – GGCTAGAGTCTTGAGGAGACCC - 3'	22	58,8	Taq	0,1	72	45''	
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.7: Condições da digestão enzimática para análise do SNP rs11886868 do gene *BCL11A* por digestão com *MboII*

		Mistura da Reação (Por Amostra)		Condições da Reação		Produto de restrição		
Gene	Variante Genética	Reagente	Volume (µL)	T (°C)	Δt	Alelo T (Normal)	Alelo C (Variante)	
<i>BCL11A</i> (cr. 2)	rs11886868	<i>MboII</i> (Thermo Scientific)	1 (5U)	37	3 horas ou <i>Overnight</i>	(+)	(-)	
		Tampão B	1,5			433 pb 226 pb	371 pb 226 pb 62 pb	
		H ₂ O	2,5	Sequência de reconhecimento				
		Produto PCR	10	5' ... G AAGA(N) ₈ ▼ ... 3' 3' ... CTTCT(N) ₇ ▲ ... 5'				

Tabela A.8: Condições de PCR para amplificação de fragmento do gene *BCL11A* para análise do rs4671393

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Ta man ho (pb)	Tm (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>BCL11A</i> (cr. 2)	rs4671393	BCL11A-4-F (Forward)	5' – CCTTCCTTTAATCAGCTTCCGCC – 3'	23	59,2	Tampão B	22,9	94	5'	533
						Primer BCL11A-4-F	0,5	94	45''	
						Primer BCL11A-4-R	0,5	58	45'' ×30	
		BCL11A-4-R (Reverse)	5'- TGGTAGCTGAGCCAAACAGATAGC – 3'	24	56,6	Taq	0,1	72	45''	
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.9: Condições de PCR para amplificação de fragmento da região intergênica *HBS1L-MYB* para análise do rs4895441

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	Tm (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>HBS1L-MYB</i> (cr. 6)	rs4895441	HBSL 1-S (<i>Forward</i>)	5'- GACAGAATCCAAGGAACCTAGGAGG- 3'	25	58,0	Tampão PCR	2,5	94	5'	778
						Primer HBSL 1-S	0,5	94	45''	
						Primer HBSL 1-AS	0,5		45'' ×30	
						BSA (0,1 M)	0,35		1'	
		HBSL 1-AS (<i>Reverse</i>)	5' – CAGTAGCAGAAAGGGAGCCG- 3'	20	57,9	dNTPS (25mM)	0,5	72		
						H ₂ O	19,55			
						Taq polimerase (5U/ µL)	0,1			
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.10: Condições da digestão enzimática para análise do rs4895441 da região intergênica *HBS1L-MYB* por digestão com *RsaI*

		Mistura da Reação (Por Amostra)		Condições da Reação		Produto de restrição		
Gene	Variante Genética	Reagente	Volume (µL)	T (°C)	Δt	Alelo A (Normal)	Alelo G (Variante)	
<i>HBS1L-MYB</i> (cr. 6)	rs4895441	<i>RsaI</i> (BioLabs)	0,5 (5U)	37	3 horas ou <i>Overnight</i>	(-)	(+)	
		Tampão CutSmart	1,5			594 pb 184 pb	494 pb 184 pb 100 pb	
		H ₂ O	3	Sequência de reconhecimento				
		Produto PCR	10	5' ... GT▼AC... 3' 3' ... CA▲TG... 5'				

Tabela A.11: Condições de PCR para amplificação de fragmento da sequência intrónica do gene *MYB* para análise do rs9399137

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	Tm (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>MYB</i> (cr. 6)	rs9399137	MYB-9399137-FW (Forward)	5' –GAAAAGCATAAGCCTGGGG- 3'	19	58	Tampão B	22,9	94	5'	307
						Primer MYB-9399137-FW (25pmol/µL)	0,5	94	45''	
						Primer MYB-9399137-R (25pmol/µL)	0,5	58	45''	
		MYB-9399137-R (Reverse)	5'-CACTGCCAGAAGCACTTTG - 3'	19	58	Taq	0,1	72		
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.12: Condições de PCR para amplificação de fragmento da sequência intrónica do gene *MYB* para análise do rs9389268

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	Tm (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>MYB</i> (cr. 6)	rs9389268	MYB-9389268-2FW (Forward)	5' –CCAGCCTGGAGTGCAGTGG- 3'	21	64	Tampão B	23,4	94	5'	344
						Primer MYB-9389268-FW (25pmol/µL)	0,5	94	45''	
						Primer MYB-9389268-R (25pmol/µL)	0,5	59	45''	
		MYB-9389268-R (Reverse)	5'-GTTTGGTGCCAAAGGTAGCAG - 3'	21	64	Taq	0,1	72		
						DNA	0,5	72	5'	
						Total	25	4	Pausa	

Tabela A.13: Condições de PCR-Longo para amplificação de todo o gene *KLF1* e região promotora para sequenciação por NGS

		Primers				Mistura de PCR		Condições PCR		
Gene	Variante Genética	Nome	Sequência	Tamanho (pb)	Tm (°C)	Reagente	Volume (µL)	T (°C)	Δt	Fragmento (Kb)
KLF1 (cr. 19)	Todo o gene e região promotora	KLF1-Ex1+Prom-FW (Forward)	5’ –CTGATAGCAGCCTCCAACGTCTGGG- 3’	25	80	FailSafe PCR Enzyme Mix	0,25	94	2’	3,2
						Primer KLF1-Ex1+Prom-FW (50pmol/µL)	0,3	94	45’’	
						Primer KLF1-Ex1+Prom-R (50pmol/µL)	0,3	63	45’’ ×35	
		KLF1-Ex3-R (Reverse)	5’- ACCTTCAGGAGCCGCTTTCT - 3’	20	62	H2O	11,15	72	4’	
						FailSafe PCR 2× PreMix D	12,5			
						DNA	0,5	72	5’	
						Total	25	4	Pausa	

Tabela A.14: Condições de PCR para amplificação do exão 1 do gene *KLF1* para sequenciação por Sanger para confirmação das variantes encontradas por NGS

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	T _m (°C)	Reagente	Volume (μL)	T (°C)	Δt	
<i>KLF1</i> (cr. 19)	Exão 1	KLF1-Ex1+Prom-FW (<i>Forward</i>)	5' –CTGATAGCAGCCTCCAACGTCTGGG- 3'	25	80	Tampão B	22,9	94	5'	735
						Primer KLF1-Ex1+Prom-FW (25pmol/μL)	0,5	94	1'	
						Primer KLF1-Ex1+Prom-R (25pmol/μL)	0,5	64	40'' ×30	
		KLF1-Ex1+Prom-R (<i>Reverse</i>)	5'- GCGTACCTCAGTCCTGGTTAAGTCTC - 3'	26	80	Taq	0,1	72	1'	
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.15: Condições de PCR para amplificação do exão 2 do gene *KLF1* para sequenciação por Sanger para confirmação das variantes encontradas por NGS

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	T _m (°C)	Reagente	Volume (μL)	T (°C)	Δt	
<i>KLF1</i> (cr. 19)	Exão 2	KLF1-Ex2B-FW (<i>Forward</i>)	5' –GGGAGGAAGAGGACGATGA- 3'	19	60	Tampão B	25	95	5'	1224
						Primer KLF1-Ex2B-FW (25pmol/μL)	0,5	95	1'	
						Primer KLF1-Ex2B-R (25pmol/μL)	0,5	58	30'' ×32	
		KLF1-Ex2B-R (<i>Reverse</i>)	5'- GGACAAGGAAGCCATAAGC- 3'	19	58	Taq AB Polimerase	0,2	72	1'30''	
						DNA	0,5	72	5'	
						Total	26,7	4	Pausa	

Tabela A.16: Condições de PCR para amplificação do exão 3 do gene *KLF1* para sequenciação por Sanger para confirmação das variantes encontradas por NGS

Gene	Variante Genética	Primers				Mistura de PCR		Condições PCR		Fragmento (pb)
		Nome	Sequência	Tamanho (pb)	Tm (°C)	Reagente	Volume (µL)	T (°C)	Δt	
<i>KLF1</i> (cr. 19)	Exão 3	KLF1-Ex3-FW (<i>Forward</i>)	5' –AGGCTGAGTAAAGGGGTGTG- 3'	20	62	Tampão B	22,9	94	5'	660
						Primer KLF1-Ex3-FW (25pmol/µL)	0,5	94	1'	
						Primer KLF1-Ex3-R (25pmol/µL)	0,5	61	40'' ×32	
		KLF1-Ex3-R (<i>Reverse</i>)	5'- ACCTTCAGGAGCCGCTTTCT - 3'	20	62	Taq	0,1	72	1'30''	
						DNA	1	72	5'	
						Total	25	4	Pausa	

Tabela A.17: Condições das reações de sequenciação de Sanger para análise dos vários SNPs estudados e para confirmação das variantes encontradas no gene *KLF1* por NGS

Gene	Variante Genética	Primer (2 pmol/ µL)	Purificação Produtos PCR		Mistura de Reação Sequenciação		Condições Reação Sequenciação	
			Reagente	Reação	Reagente	Volume (µL)	T (°C)	Δt
<i>BCL11A</i> (cr. 2)	rs4671393	BCL11A-4-R	5 µL Produto PCR + 2 µL Exosap-IT	37°C 15'	Tampão Big Dye (5×) Big Dye (2×) Primer (2 pmol/ µL) H ₂ O DNA Purificado	1,75	96	4'
<i>MYB</i> (cr. 6)	rs9399137	MYB-9399137-FW		80°C 15'		0,5	96	10''
	rs9389268	MYB-9389268-FW				1	55	5'' ×25
<i>KLF1</i> (cr.19)	Exão 1	EX1+Prom-FW + Ex1+Prom-R		4°C Pausa		5,25	60	4'
	Exão 2	Ex2B-FW + Ex2B-R				1,5	60	8'
	Exão 3	Ex3-FW + Ex3-R				Total: 10	4	Pausa